GENOMIC DISCOVERY OF RECURRENT *CD44-SLC1A2* GENE FUSION IN GASTRIC CANCER

TAO JIONG
B.Sc (Honors), NUS

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSIOLOGY

YONG LOO LIN SCHOOL OF MEDICINE

NATIONAL UNIVERSITY OF SINGAPORE

2011

ACKNOWLEDGEMENT

I am extremely grateful and indebted to my supervisor Associate Professor Patrick Tan and my co-supervisor Dr. Celestial Yap for their advice and invaluable guidance and encouragements throughout the course of my research. Without their support and encouragement, this work would not have been accomplished. It was an eye-opening and wonderful experience to conduct research under their supervision.

I will also like to express my earnest thanks to Dr Hue Kian Oh, for her helpful suggestions to my project, unwavering support as well as expert opinions. Her intellectual contribution and logical thinking process have enhanced my knowledge which has been of great value to me.

I would also like to thank Nian Tao Deng for his bioinformatics support to this project; the gratitude also goes out to Dr. Baohua Huang, Dr. Iain Beehuat Tan, Dr. Chia Huey Ooi, Jeanie Wu, Minghui Lee, Shenli Zhang for their contributions and involvement in this project.

I would like to thank Dr. Siew Hong Leong and Prof. Oi Lian Kon from National Cancer Cencer Singapore for their work on Spectral Karyotyping; Seong Soo Lim and Dr. Valere Cacheux from Genome Institute of Singapore for their work on Fluorescent *in situ* Hybridization (FISH) and Fiber-FISH; I would like to extend my appreciation towards Dr. Kalpana Ramnarayanan, who did the Array-CGH profiling of cell lines and tumor samples. Without her great effort, I simply cannot start my project. I would like to acknowledge Prof. Nallasivam Palanisamy from University of Michigan. Prof. Nallasivam Palanisamy gave us lots of valuable suggestions and comments on our work.

I am grateful to Prof. Sun Young Rha, Prof. Hyun Cheol Chung, Prof. Duane T. Smoot and Prof. Hassan Ashktorab for their generous gifts of cell lines.

Special thanks belong to my lab colleagues who have given me excellent cooperation and assistance throughout my stay in the lab. I am honored to have had the opportunity to work with each and every one of them in different aspects of my research. In them, I have found firm friends and I truly cherish the friendship we share.

The work was supported by Biomedical Research Council BMRC 05/1/31/19/423, National Medical Research Council NMRC TCR/001/2007, and institutional funding from Duke-NUS and Cancer Sciences Institute of Singapore. I wish to acknowledge Department of Physiology, Yong Loo Lin School of Medicine in National University of Singapore, for providing various supports from course education to the management of student affair.

I wish to acknowledge my deepest gratitude and appreciation to my family, especially my husband, who has been my constant source of encouragement and moral support, my pillar of strength and my confidante, without whom this journey would have been that much harder, and I dedicate this thesis to him.

TABLE OF CONTENTS

SUMMARY

Gastric cancer (GC) is the second leading cause of cancer death worldwide. Despite declining incidence rates globally, the overall five year survival rate of GC is less than 24%, which is much lower compared to other cancers. Early stage stomach cancer is often difficult to diagnose because of nonspecific symptoms. Therefore, understanding the pathogenesis and biological features, as well as identification of new markers and therapeutic targets of GC are crucial to improve its detection and therapy.

Like many other cancers, chromosomal instability is frequently observed in GC. Detailed characterization of the aberrant regions in cancer has identified several potential oncogenes and tumor suppressor genes that may contribute to carcinogenesis. Among the various genomic abnormalities associated with cancers, fusion genes and transcripts are particularly notable due to their cancer-specific nature and their translational potential as diagnostic and therapeutic targets. Although previously largely restricted to hematologic malignancies, recent studies have shown that fusion genes in solid epithelial tumors can also be elucidated using high-resolution genomic approaches. For example, *TMPRSS2-ERG* was identified in prostate cancer and *EML4-ALK* in non-small-cell lung cancer.

Therefore, using detailed fine-scale survey of genomic copy number alterations (CNAs), our objective is to identify possible fusion transcripts in GC which may provide further mechanistic insights into GC development and highlight opportunities for early detection and new therapies.

We profiled a discovery cohort of 133 GCs (106 primary tumors and 27 cell lines) using high density array-based comparative genomic hybridization (aCGH) microarrays. To nominate potential fusion genes, we used a technique called genomic breakpoint analysis (GBA), previously used to identify fusion genes in leukemia. With this strategy, we discovered several tumors exhibiting recurrent genomic breakpoints in the *SLC1A2/EAAT2 gene*, encoding a glutamate transporter. Subsequent 5' RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) analysis of a GC cell line with *SLC1A2* breakpoints (SNU16) revealed the expression of a *CD44-SLC1A2* fusion transcript caused by a paracentric chromosomal inversion, which produced a truncated but functional SLC1A2 protein. Using custom-designed fusion-specific siRNAs, we showed that silencing of *CD44-SLC1A2* in fusion-positive SNU16 cells significantly reduced cellular proliferation, invasion, and colony formation, but not in cell lines lacking *CD44-SLC1A2* expression. Conversely, *CD44-SLC1A2* overexpression in gastric cells stimulated these pro-oncogenic traits. In addition, *CD44-SLC1A2* silencing also significantly reduced intracellular glutamate levels and sensitized SNU16 cells to cisplatin, a commonly used chemotherapeutic agent in GC.

We further demonstrated that recurrent *CD44-SLC1A2* fusions were observed in primary gastric tumors. Although *CD44-SLC1A2* expression was relatively rare in unselected GCs (2/43), the percentage of tumors testing positive for the *CD44-SLC1A2* fusion was markedly increased when we screened GCs preselected for high *SLC1A2* expression. In addition, we also found that genes upregulated in *SLC1A2*-high expressing tumors were significantly enriched in genes related to ribosomal function and protein translation,

supporting the notion that *SLC1A2*-high expressing tumors comprise a distinct molecular subclass.

In conclusion, our study contributes not only to the identification of a recurrent novel fusion gene candidate in GC, which contributes to GC development and plays a role in cancer metabolism, but also suggests *CD44-SLC1A2* may be a potential diagnostic marker and therapeutic target in GC.

LIST OF TABLES

Results

LIST OF FIGURES

Introduction

Material and methods

Results

ABBREVIATIONS

ABL1	Abelson murine leukemia viral oncogene homolog 1
aCGH	Array-based comparative genomic hybridization
AICR	American Institute for Cancer Research
ALK	Anaplastic lymphoma kinase
ALL	Acute lymphoblastic leukemia
AML1	Acute myeloid leukemia 1
APL	Acute promylocytic leukemia
ATP	Adenosine triphosphate
BCR	Breakpoint cluster region
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BTAK	Aurora kinase A
CCNE1	Cyclin E1
CDH1	E-cadherin
CGH	Comparative genomic hybridization
CML	Chronic myelogenous leukemia
CNA	Copy number aberration
CNS	Central nervous system
CSC	Cancer stem cell
DMEM	Dulbecco's Modified Eagle Medium
EAAT	Excitatory amino acid transporter
ECM	Extra cellular matrix
EDTA	Ethylenediaminetetraacetic acid
EML4	Echinoderm microtubule associated protein like 4
EPIC	European Prospective Investigation into Cancer and Nutrion
ERBB	Epidermal growth factor receptor
ERG	v-ets erythroblastosis virus E26 oncogene homolog
FBS	Fetal bovine serum
FISH	Fluorescence *in situ* Hybridization
GBA	Genomic breakpoint analysis
GC	Gastric cancer
GCCL	Gastric cancer cell line
H. Pylori	Helicobacter Pylori
HR	Hazard ratio
MYC	v-myc myelocytomatosis viral oncogene homolog
P53	Tumor protein p53
PAX5	Paired box gene 5
PBS	Phosphate buffered saline
PML	Promyelocytic leukemia
RARA	Retinoic acid receptor, alpha
RIPA	Radioimmunoprecipitation
RLM-RACE	RNA-ligase mediated rapid amplification of cDNA ends
RPMI	Roswell park memorial institute

RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcription-polymerase chain reaction
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SKY	Spectral karyotyping
SLC	Solute carrier protein
TCA	tricarboxylic acid
TCR	T cell receptor
TEL	Telomere elongation
TEMED	N,N,N',N'-Tetramethylethylenediamine
TMPRSS2	Transmembrane protease, serine 2
WCRF	World Cancer Research Fund

LIST OF PUBLICATIONS

1. **Tao J**, Deng NT, Ramnarayanan K, Huang B, Oh HK, Leong SH, Lim SS, Tan IB, Ooi CH, Wu J, Lee M, Zhang S, Rha SY, Chung HC, Smoot DT, Ashktorab H, Kon OL, Cacheux V, Yap C, Palanisamy N, Tan P. CD44-SLC1A2 Gene Fusions in Gastric Cancer. *Sci Transl Med.* 3(77):77ra30 (2011).

2. Zang ZJ, Ong CK, Cutcutache I, Yu W, Zhang SL, Huang D, Ler LD, Dykema K, Gan A, **Tao J**, Lim S, Liu Y, Futreal PA, Grabsch H, Furge KA, Goh LK, Rozen S, Teh BT, Tan P. Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing. *Cancer Res.* 71(1):29-39 (2011).

3. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, **Tao J**, Ward L, Koo JH, Gopalakrishnan V, Zhu Y, Cheng LL, Lee J, Rha SY, Chung HC, Ganesan K, So J, Soo KC, Lim D, Chan WH, Wong WK, Bowtell D, Yeoh KG, Grabsch H, Boussioutas A, Tan P. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLoS Genet.* 5(10):e1000676 (2009).

4. Hou Q, Wu YH, Grabsch H, Zhu Y, Leong SH, Ganesan K, Cross D, Tan LK, **Tao J**, Gopalakrishnan V, Tang BL, Kon OL, Tan P. Integrative genomics identifies RAB23 as an invasion mediator gene in diffuse-type gastric cancer. *Cancer Res.* 68(12):4623-30 (2008).

5. ” **Fusion Genes in Gastrointestinal Cancer” has been filed for US patent on 5, April 2011.**

Abstract: “Genomic Discovery of *CD44-SLC1A2* Gene Fusions in Gastric Cancer”

Been accepted and selected as the only Asian speaker to present research projects at the American Association Cancer Research (AACR) Conference, Translational Cancer Medicine July, 2010 (CA, USA)

Chapter I: Introduction

1.1 Gastric cancer

1.1.1 Introduction

Gastric adenocarcinoma, or gastric cancer (GC) is a very common disease worldwide and the second most frequent cause of cancer death, affecting about one million people per year (Hartgrink et al., 2009). The geographical distribution of stomach cancer is characterized by wide international variations (Rastogi et al., 2004) (Fig. 1.1). The incidence is particularly high in Northeast Asia (Japan, Korea and China), intermediate incidences in Eastern Europe, and parts of Central and South America. Incidences are low in Southern Asia, North and East Africa, North America, Australia, and New Zealand (Brenner et al., 2009; Hartgrink et al., 2009; Parkin et al., 2005). In Singapore, it ranks as the 4th most prevalent cancer in males and 6th most common cancer in females (Lim et al., 2009). Similar with most other solid tumors, the incidence of GC increases with age and the cancer is relatively rare in male or female patients younger than 45 years. Most patients are between 60 and 80 years old at diagnosis. In general, incidence and mortality rates in men are approximately double to those in women.

Despite the declining incidence globally, the overall five year survival rate of GC is less than 24%, which is much lower compared to the other cancers such as breast cancer and colorectal cancer (five year survival rates 88.7% and 64.4% respectively) (LAG et al., 2008). Particularly prevalent in several Asian countries (Kamangar et al., 2006), most GC patients present with advanced stage disease with the exception of patients in Japan and Korea where early detection screening programs are practiced using either barium

photofluorography or endoscopy (Hamashima et al., 2008; Lee, 2006). Current strategies for treating GC patients are far from optimal, with conventional surgery and chemotherapy regimens conferring modest survival benefits and median survival times of 7-10 months (Jackson et al., 2009) .

1.1.2 Histological subtypes

Gastric carcinomas occur through successive changes. According to Lauren's Classification, which is the most widely used and accepted approach to classify GC, GC is classified into two distinct subtypes: (1) the well differentiated or intestinal type; (2) the undifferentiated or diffuse type (Krejs, 2010; Lauren, 1965).

The development of the intestinal type GC includes the transformation of the normal mucosa into a mucosa that resembles intestinal epithelium (intestinal metaplasia). The presence of intestinal metaplasia increases the risk of GC. Subsequently, intestinal metaplasia may progress to dysplasia, and ultimately to carcinoma (Stemmermann, 1994). By contrast, diffuse type GC presumably arises as single-cell changes in the mucus neck region of the gastric glands. These cells may subsequently proliferate and invade out from the crypt (Fig. 1.2). However, some gastric cancers failed to be classified into either subtype as they present mixed features of diffuse and intestinal subtypes. These tumors are thus usually regarded as a mixed subtype.

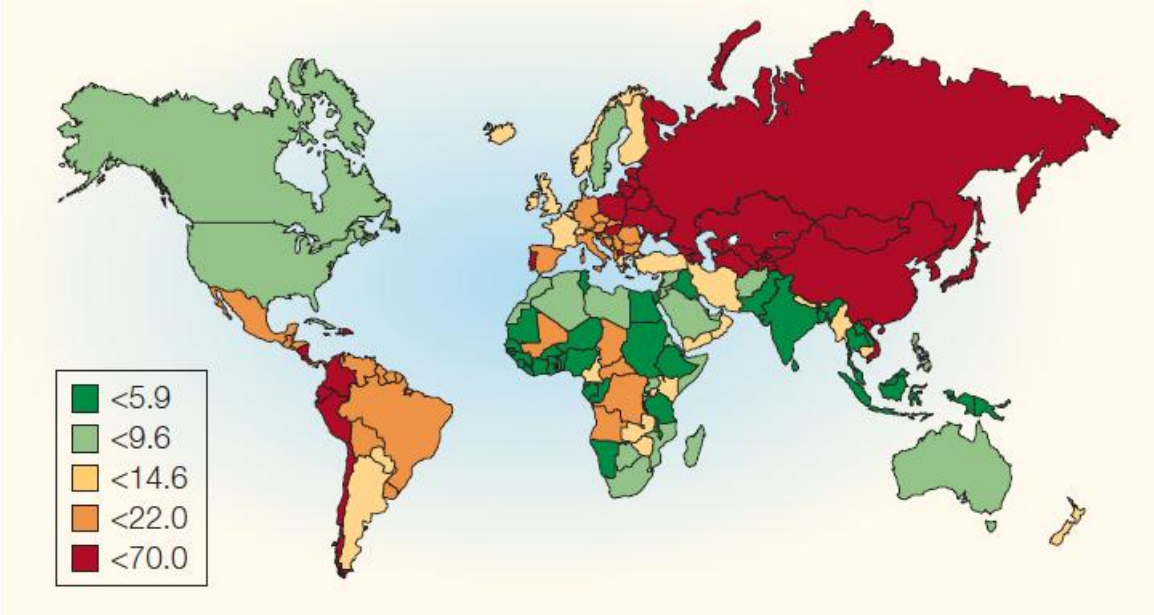


Figure 1.1. Global variations in GC incidence
The incidence of stomach cancer for men of all ages is highest (orange and red) in developing countries such as Asia and South America, and lowest (light and dark green) in North America, parts of Africa, India and Australia.

Adapted from Nat Rev Cancer *4*, 909-917. (Rastogi et al., 2004)

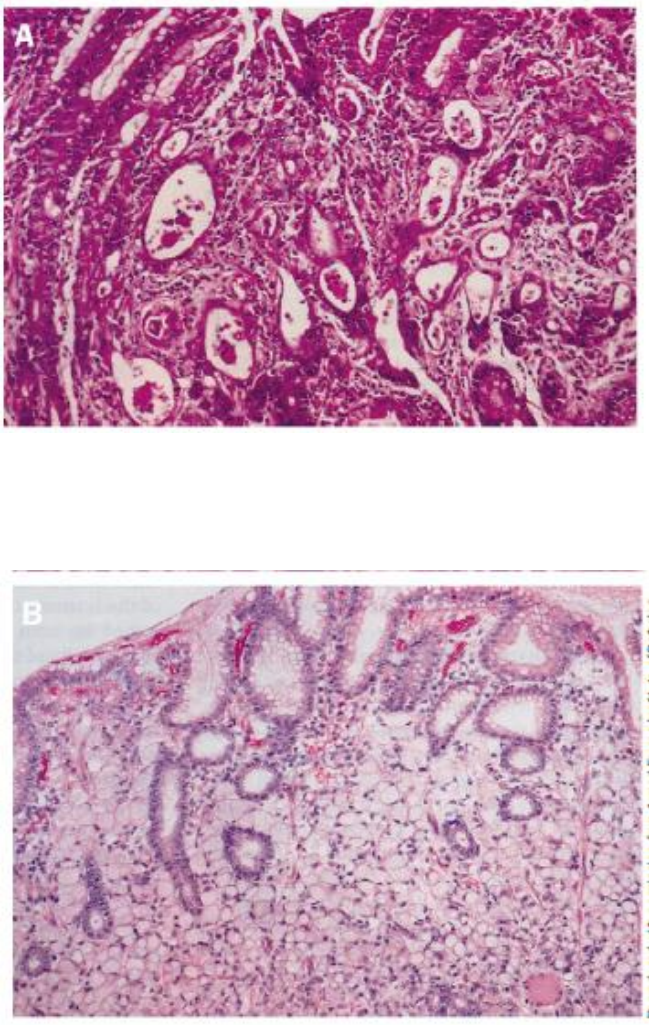


Figure 1.2. Histological subtypes of GC
Upper, gastric carcinoma of intestinal type. Normal mucosa is replaced by infiltrating tubular profiles.
Lower, signet-ring cell gastric carcinoma of the diffuse type. There is diffuse infiltration of the mucosa by signet ring cells. The gastric pits appear normal and there is no evidence of gastritis or intestinal metaplasia.

Adapted from Lancet *362*, 305-315. (Hohenberger and Gretschel, 2003)

1.1.3 Risk factors

Clinical risk factors for GC include *Helicobacter pylori* (*H. pylori*) infection, high intake of various traditional salt-preserved foods and salt, low consumption of fresh fruits and vegetables, smoking and genetic polymorphisms of the host (Hartgrink et al., 2009). While familial patterns of GC incidence have been reported, most GC cases are sporadic (Barber et al., 2006).

In 1984, Marshall and Warren redetected a gram-negative bacillus, now called *H. pylori*, which is a common inhabitant of the human stomach (Marshall and Warren, 1984). Numerous studies conducted since then have consistently found *H. pylori* infection to be related to an increased risk of stomach cancer. In 1994, *H. pylori* was classified as a definite carcinogen for GC (International Agency for Research on Cancer, 1994). Infection is usually acquired in childhood. Although the exact mechanisms of action in the complex process of stomach cancer are unknown, its action is probably indirect by provoking gastritis, a precursor of gastric atrophy, metaplasia, and dysplasia. The association between the infection and the subsequent risk of developing GC is about six-fold (Group, 2001). However, it has been reported that about half of the adult world population has been infected but only a small proportion of infected subjects will eventually develop GC (Rothenbacher and Brenner, 2003). This indicates that there may be other contributing factors involved.

Diet plays an important role. Recently, the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR) did an extensive report on the scientific literature on diet and physical activity on the prevention of cancer. They

concluded that GC is mostly preventable by appropriate diets and associated factors (WCRF and AICR, 2007). Risk is increased by high intake of salt and some traditionally preserved salted foods, especially meats and pickles (Palli, 2000).

Additionally, environmental factors have also been recognized as potential factors related to the risk of GC. Migrant populations from high-risk countries to low-prevalence countries show a significant reduced risk in their second generation. The data fit with observations concerning the importance of childhood environment in determining the risk (Coggon et al., 1990).

Smoking is another risk factor for GC. The European Prospective Investigation into Cancer and Nutrition (EPIC) project found a significant association between cigarette smoking and GC risk: the hazard ratio (HR) was 1.45, 1.7 and 1.8 for ever smokers, current male and current female smokers respectively. The HR increased with intensity and duration of cigarette smoking. In addition, approximately 18% of GC may be attributable to tobacco smoking.

Host factors, such as polymorphisms in cytokine genes (eg. Interleukin 1β, interferon γ receptor 1) are associated with the increased progression risks of GC (Canedo et al., 2008; El-Omar et al., 2000; Fox et al., 2000; Hold et al., 2007).

1.1.4 Prevention and early detection

Patients presenting clinical manifestations of GC have limited options for cure. Thus, early detection and prevention play a key role in the fight against GC.

Several studies have gained evidence that the eradication of *H. pylori* is one of the most promising preventive strategies in the fight against GC (Fry et al., 2007; Malfertheiner et al., 2005). They have demonstrated that the earlier *H. pylori* gets eradicated, the lower risk of getting GC (Fukase et al., 2008; Wu et al., 2009). However, some other studies showed no benefit of *H. pylori* eradication. One research group reported two cases of GC development 4 and 14 years after *H. pylori* eradication (Cannizzaro and De Paoli, 2009; de Vries et al., 2009). These patients presented with gastric ulcer and preneoplastic changes (i.e. IM and gastric atrophy) at baseline. Though the role of eradication as the main preventive strategy continues to be questioned (De Vries and Kuipers, 2007), the effect of eradication and the subsequent risk of developing GC depend on the degree and extent of preneoplastic changes (i.e. gastric atrophy and intestinal metaplasia) at the time of eradication (Selgrad et al., 2010) .

This suggests that eradication of *H. Pylori* does not prevent GC in all cases, especially those already present with preneoplastic changes. Therefore, it seems that the earlier the bacterium gets eradicated, the more significant is the decrease of GC risk. In addition, improved knowledge of molecular changes in precursor lesions might enable further discrimination between patients at high and low risks. Studies could establish which patients will benefit from *H. pylori* eradication.

Besides *H. pylori* eradication, the dominant dietary hypothesis is that fresh fruits and vegetables, or contained micronutrients, are protective against GC. Numerous studies have shown a protective association with fresh fruits and vegetables, independent of other dietary factors (Block et al., 1992; Kono and Hirohata, 1996). Possible protective micronutrients include vitamins C (ascorbate) and E (alpha-tocopherol), carotenoids (particularly beta carotene), and selenium (Kono and Hirohata, 1996). The evidence is strongest for vitamin C, with an approximate halving of risk associated with high intake vs. low intake demonstrated in case–control studies (Neugut et al., 1996).

Population-based screening represents one of the best options for the primary prevention and early detection of GC especially in the high-risk regions. In Japan, photofluorography has been used for GC screening since 1960. About 6 million people are screened annually (Miki, 2006). Serum pepsinogen test, a new and potentially useful method, was introduced for mass screening to identify individual with atrophic gastritis who are at high risks for GC (Miki, 2006). To predict the risk of GC development and to diagnose atrophic gastritis, serologic testing for a combination of pepsinogen (PG) I and II, and gastrin and *H. pylori* antibodies have yielded accurate results over the last years (Selgrad et al., 2010; Vaananen et al., 2003; Watabe et al., 2005).

1.1.5 Treatment

Current strategies for treating GC patients are far from optimal, with conventional surgery and chemotherapy regimens conferring modest survival benefits and median survival times of 7 to 10 months (Jackson et al., 2009).

Surgical resection of the primary tumor and regional lymph nodes is the treatment of choice for GC. However, the extent of disease, the operative procedure, and patient selection are crucial in optimizing outcome. Adjuvant therapy (mainly, chemotherapy±radiotherapy) still warrants further evaluation for high-risk GC patients. Neoadjuvant therapy may reduce tumor mass enabling resection with potentially curative intent (Catalano et al., 2009).

Since most GC patients are diagnosed at advanced stages where the tumors are unresectable, systemic chemotherapy is the main treatment option. Many single agents and combinations are active in the treatment of the metastatic disease. The commonly used chemotherapeutic agents include platinum compounds (such as cisplatin), fluoropyrimidines, anthracyclines, and, recently, taxanes and irinotecan. The objective response rates range from 10% to 30% for single-agent therapy and 30% to 60% for combination regimens (Sastre et al., 2006). Although a large number of chemotherapy regimens have been tested in randomized studies, there is no internationally accepted standard of care, and uncertainty remains regarding the choice of the chemotherapy regimen (Wagner et al., 2006).

1.1.6 Genetic and genomic alterations in GC

Like many other cancers, previous cytogenetic studies of GC karyotypes have revealed a complex portrait of recurring patterns of chromosomal amplifications and deletions associated with GC, including signature genomic gains at chromosomes 7q, 8p, 17q, 20q and losses in 5q, 9p, 18q (Kang et al., 2006; Kimura et al., 2004; Nessling et al., 1998; Sakakura et al., 1999; Tay et al., 2003). The recurrent nature of these aberrations has been attributed to the presence of genes important for gastric carcinogenesis, such as *CD44* at 11p13, *CCNE1* at 19q13, and *BTAK* at 20q13 (Benusiglio et al., 2006; Fukuda et al., 2000; Husdal et al., 2006; Sen et al., 1997; Tahara, 1995). However, identifying specific oncogenes and tumor suppressor genes within such regions is often challenged by their large sizes covering tens to hundreds of genes. For example, chromosome 6p amplifications have been reported in GC at frequencies ranging from 2-4% using conventional comparative genomic hybridization (CGH) platforms to 85% (6p22) by cDNA based aCGH (Boussioutas et al., 2003; Gorringe et al., 2005; Nakanishi et al., 2000; Sakakura et al., 1999), but the specific genes representing the targets of amplification in the 6p region are currently unknown.

Studies investigating the genetic basis of GC have also identified germline polymorphisms in cytokine genes (e.g. interleukin 1β, TNF-α) (El-Omar et al., 2000; Hold et al., 2007). IL-1β and TNF-α are pro-inflammatory cytokines and acid inhibitors highly expressed in *H. pylori*-induced gastritis. Host genetic polymorphisms that affect IL-1β and TNF-α may associate with increased risk of developing GC.

In approximately 30% of familial gastric cancers, a germline mutation in the E-cadherin gene (*CDH1*) is identified (Guilford et al., 1998). *CDH1* encodes a transmembrane cellular adhesion protein acting as a mediator of homophilic recognition signals, leading to cell-cell contact inhibition. The majority of *CDH1* mutation carriers were considered susceptible to this inherited cancer syndrome dominated by diffuse GC, suggesting a central role for this gene as a tumor suppressor in diffuse GC (Guilford et al., 1999).

Somatic mutations in tumor suppressor genes such as *p53* and *RUNX3* have been reported in GC (Li et al., 2002; Pan et al., 2008; Tamura et al., 1991). The tumor protein 53 gene (*p53*), the most frequently studied tumor suppressor gene, plays multiple roles in carcinogenesis in response to cellular stresses including apoptosis, cell cycle arrest, senescence, DNA repair, cell metabolism or autophagy by activating specific target genes (Kruse and Gu, 2009). *p53* is the most frequently mutated gene in human cancers. In GC, around 45% cases carry the mutated *p53* (Soussi et al., 2000).

RUNX3 belongs to the *RUNX* family of transcriptional factors with conserved Runt-DNA binding domain (Chuang and Ito, 2010). Loss of *RUNX3* is increasing with tumor grade in primary GC samples. Between 45% and 60% of human GC cells do not significantly express *RUNX3* due to hemizygous deletion and hypermethylation of the *RUNX3* promoter region (Li et al., 2002).

Although knowledge of genomic aberrations and potential oncogenes and tumour-suppressor genes associated with GC are regarded as a means to reveal new therapeutic targets or predictive markers of therapy response, no such biomarker is yet available.

Therefore, identifying additional molecular aberrations in GC may provide further mechanistic insights into GC pathogenesis and highlight opportunities for early detection and novel therapies.

1.2 Fusion gene

1.2.1 Introduction

A substantial body of cytogenetic and molecular work over the past has established that aneuploidy is a hallmark feature of many cancers, with individual tumors frequently exhibiting aberrant patterns of genomic amplifications, deletions, and translocations (Albertson et al., 2003; Kallioniemi, 2008). Among the various genomic aberrations, fusion genes make an important contribution to the development of cancer due to their cancer-specific nature (Heim and Mitelman, 2008).

Fusion genes are hybrid genes formed by the combination of two previously separate genes (Fig.1.3). In cancers, fusion genes can be produced by genomic amplifications, deletions, translocations and rearrangements (Mitelman et al., 2007). The prevailing view is that such fusion transcripts and proteins are abnormal thereby providing tumor cells with growth and survival advantages. Most importantly, they may serve as specific diagnostic markers and therapeutic targets due to their cancer specific nature.

The first specific translocation identified by Nowell and Hungerford in 1961 in human neoplasia was t(9;22)(q34;q11), resulting in the Philadelphia chromosome (Nowell and Hungerford, 1960) (Fig.1.4). The fusion gene brings together the 5' part of *BCR* gene at 22q11 and the 3' part of the *ABL1* tyrosine kinase-encoding gene at 9q34, leading to a hybrid BCR-ABL1 protein with increased tyrosine kinase activity (Look, 1997; Rabbitts, 1994), which is essential for the initiation, maintenance and progression of Chronic Myelogenous Leukemia (CML) (Ren, 2005). *BCR–ABL1* is a diagnostic marker for CML. Detection of *BCR–ABL1* fusion transcript in peripheral blood is used to confirm CML

diagnosis, and to monitor cytogenetic remission and residual disease (Ren, 2005). Most importantly, *BCR–ABL1* is a specific therapeutic target. Imatinib (or Gleevec), a small-molecule inhibitor of ABL1 tyrosine kinase activity, is used as the standard treatment for chronic-phase CML (Druker et al., 2006; Hughes et al., 2003; O'Brien et al., 2003). Subsequently, the identification of the t(8;14)(q24;32) in Burkitt's lymphoma, which juxtaposes *MYC* with the immunoglobulin heavy chain gene, dramatically increased our understanding of the pathogenetic significance of translocations and gene fusions in the origin of human cancers (Rabbitts and Boehm, 1991). Since then, an increasing number of gene fusions have been recognized as important diagnostic and prognostic markers in malignant haematological disorders and sarcomas. The biological and clinical impact of gene fusions in the more common solid tumor types are less appreciated. Until recently, with the advent of new and powerful investigative tools, fusion genes have been identified in solid tumors, suggesting that causal gene rearrangements exist in common epithelial cancers.

1.2.2 Types of gene fusions

There are generally two types of gene fusions. In the first, the promoter or the enhancer element of one gene is juxtaposed to an oncogenic gene, leading to an up-regulation of the second gene (eg *IgH-Myc*) (Fig.1.5) (Rabbitts and Boehm, 1991). As a consequence of the translocations, the *MYC* gene becomes constitutively expressed owing to the influence of regulatory elements of the immunoglobulin genes. An analogous scenario is seen in T-cell lymphomas and T-cell acute lymphoblastic leukemia (ALL) in which regulatory elements of T-cell receptor genes deregulate the expression of various 3′

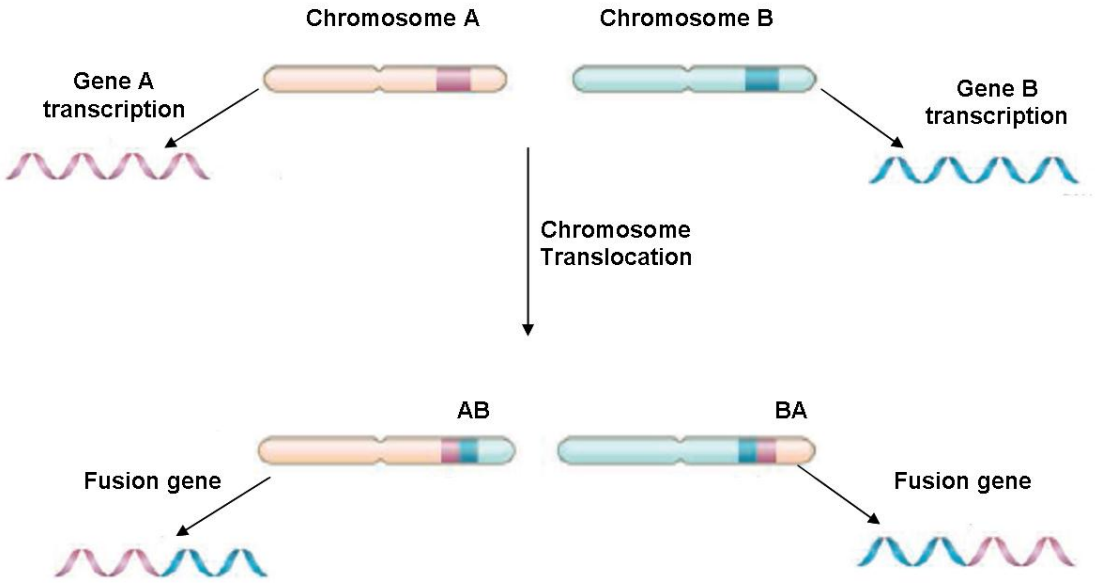


Figure 1.3. Fusion RNAs
Chromosomal translocation can give rise to fusion mRNAs. Some chromosomal translocations produce two hybrid genes that may produce mRNAs containing the 5' end of one gene and the 3' end of the other.

Modified from Science *32*, 1302-1304.(Rowley and Blumenthal, 2008)

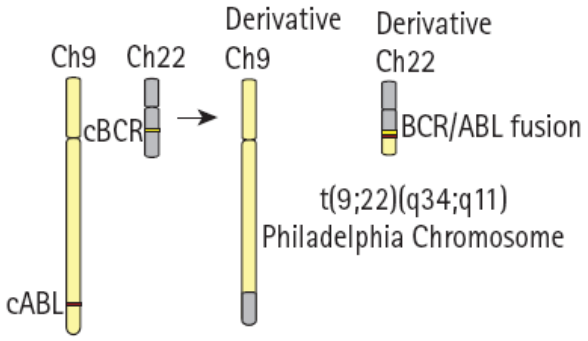


Figure 1.4. Philadelphia chromosome
The BCR-ABL fusion protein results from reciprocal translocation of the *BCR* gene from chromosome 22 with the *ABL* gene from chromosome 9.

Adopted from BJU Int *102*, 276-282. (Morris et al., 2008)

partner genes (Aplan, 2006; Mitelman et al., 2007). Immunoglobulin and T-cell receptor genes are frequently involved in chromosomal aberrations because they are naturally rearranged to generate active antigen-receptor genes. This process occasionally, in error, leads to an interchromosomal translocation or inversion. In addition, deregulation as a consequence of promoter swapping or substitution has also been identified in some benign and malignant solid tumors (eg. *TMPRSS2-ETS* in prostate cancer) (Tomlins et al., 2005).

In the second type, gene fusions lead to a production of chimeric protein with oncogenic signaling potential (eg *BCR-ABL*) (Fig.1.6) (de Klein et al., 1982). The fusion of *BCR* sequences to *ABL* generates a new protein with increased tyrosine kinase activity of ABL, and brings new regulatory domains/motifs to ABL. Many hybrid genes, accounting for approximately 75% of the known gene fusions in malignant disorders, have been identified in various haematological malignancies and solid tumors (Mitelman et al., 2007).

To date, most of reported fusion events involve two main groups — transcription factors and tyrosine kinases. They account for 50% of the genes involved and distributed quite equally among haematological disorders and solid tumors (Mitelman et al., 2004).

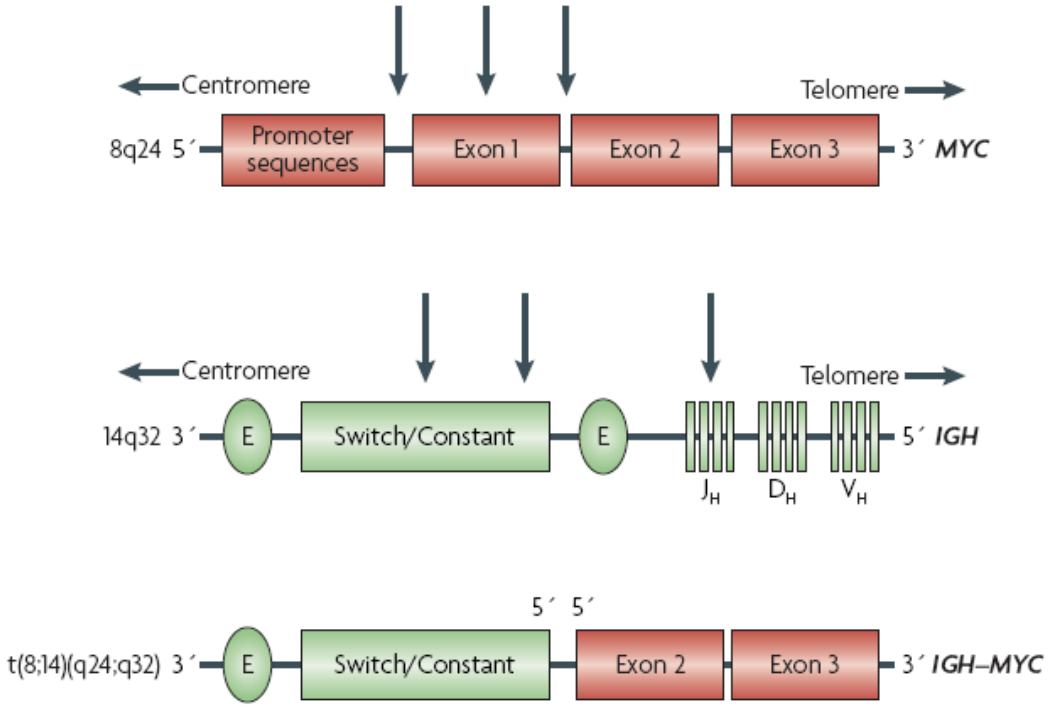


Figure 1.5. Gene fusion leading to gene upregulation (Type 1)
The t(8;14)(q24;q32) translocation, the most common translocation in Burkitt lymphoma, leads to the deregulation of the *MYC* gene at 8q24 through its juxtaposition with regulatory elements of the immunoglobulin heavy chain (*IGH*) gene at 14q32; that is, the *MYC* gene is constitutively activated because its expression is driven by immunoglobulin enhancers (E).

Adapted from Nat Rev Cancer *7*, 233-245. (Mitelman et al., 2007)

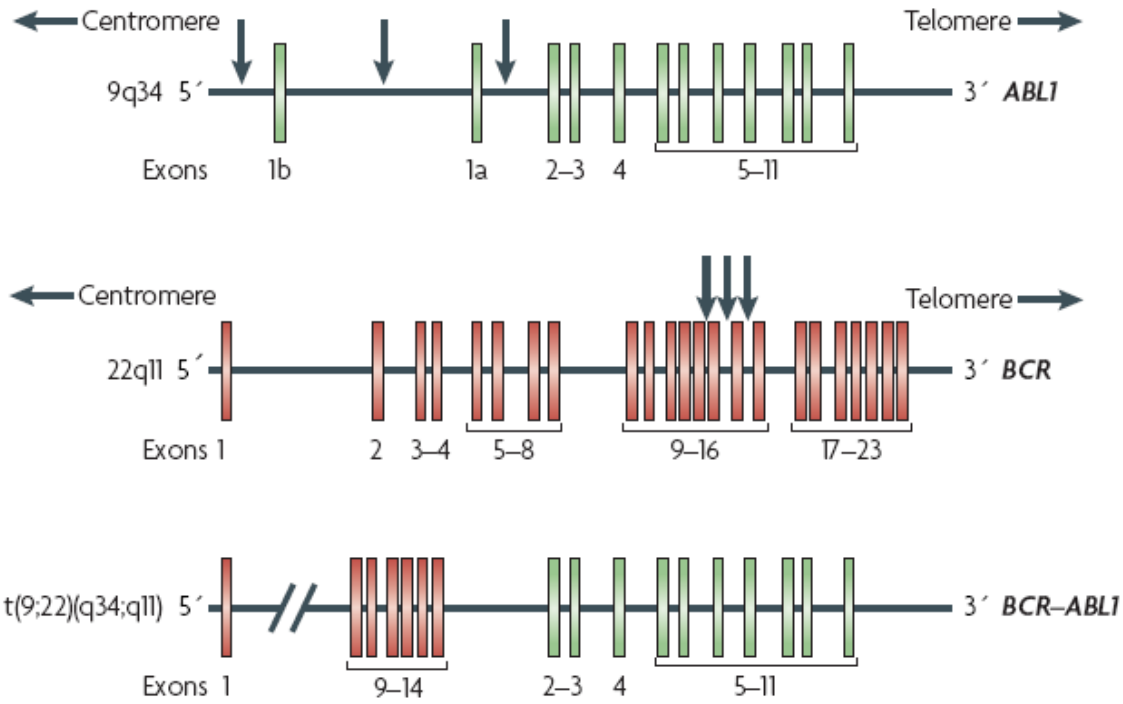


Figure 1.6. Gene fusion leading to a chimeric gene (Type 2)
The Philadelphia chromosome, which originates through the translocation (9;22)(q34;q11), juxtaposes the 5′ part of the *BCR* gene at 22q11 with the 3′ part of the *ABL1* gene at 9q34, resulting in the creation of a hybrid *BCR–ABL1* fusion gene.

Adapted from Nat Rev Cancer *7*, 233-245. (Mitelman et al., 2007)

1.2.3 Balanced and unbalanced rearrangements

Gene fusions can arise from either balanced or unbalanced rearrangements. Balanced rearrangements refer to chromosome abnormalities that give rise to structurally altered chromosomes without the gain or loss of genetic material. Such changes comprise of reciprocal translocations and inversions. On the other hand, unbalanced rearrangements refer to chromosomal abnormalities involving gain or loss of genetic material, such as duplications, amplification insertions or deletions.

To date, more than 300 fusion genes have been identified in human malignancies, the majority of which are the result of balanced chromosomal rearrangements (Mitelman et al., 2007). Examples of fusions resulting from balanced translocations are *IGH–MYC* in Burkitt lymphoma/leukaemia rabbitts (Rabbitts and Boehm, 1991); *BCR–ABL1* in CML (de Klein et al., 1982); *PML–RARA* in acute promyelocytic leukaemia (Ghaffari et al., 2006); *JAZF1–SUZ12* in endometrial stromal sarcoma.

Examples of gene fusions that are typically detected in the context of an unbalanced cytogenetic rearrangement include *COL1A1–PDGFB* in dermatofibrosarcoma protuberans with supernumerary ring chromosomes (Sirvent et al., 2003); *NUP214–ABL1* (Graux et al., 2004) occurring on amplified episomes in T-cell ALL; *TMPRSS2–ERG* in prostate cancer (Tomlins et al., 2005). The majority of translocations described in human cancer are unbalanced (Mitelman et al., 2004). The analysis of unbalanced translocations has largely failed to identify target genes because of the heterogeneity of the chromosomal breakpoints and the multiplicity of partner chromosomes. Thus, it has been assumed that these rearrangements affect gene function through the loss or gain of

chromosomal material. Identification of the key molecular events resulting from unbalanced rearrangements would be a significant step toward understanding their role in cancer pathogenesis (An et al., 2008).

1.2.4 Fusion genes in hematological malignancies

To date, more than 80% of gene fusions have been identified in haematological disorders, including malignant lymphomas and most of them were identified through standard cytogenetic assays such as spectral karyotyping (SKY) or fluorescence *in situ* (FISH) Hybridization (Maher et al., 2009; Mitelman et al., 2007). Fusion events have been well documented and recognized as causal events to some heamatological malignancies. They have been used to identify particular cancer subtypes (eg *PML-RAR* in acute promyelocytic leukemia) and specifically targeted using drugs (eg *BCR-ABL* in CML) (Deininger et al., 2005; Ghaffari et al., 2006).

The acute promyelocytic leukemia (APL)-specific t(15;17) chromosome abnormality is characterized at the molecular level by rearrangement of the *PML* and *RAR alpha* genes, resulting in fusion *PML/RAR alpha* mRNA and a chimeric protein. Besides its relevance in the pathogenesis of the disease, this hybrid gene represents a specific tumor marker. Several studies have highlighted the clinical relevance of *PML/RAR alpha* detection to identify APL. In fact, this hybrid gene can be detected in 100% of APLs (Ghaffari et al., 2006).

In CML, the most frequent chromosomal translocation is t(9;22) which causes fusion of the BCR signaling protein to the ABL non-receptor tyrosine kinase, resulting in

constitutive tyrosine kinase activity and complex interactions of this fusion protein with many other transforming elements, such as the signaling pathway for RAS (GTP-binding protein that activates target genes involved in cell differentiation, proliferation, and survival) (Ren, 2005). As an activated kinase, BCR-ABL offers an attractive therapeutic target, and imatinib, a small-molecule inhibitor of the ABL kinase, has proven effective against leukaemias that express *BCR-ABL* (Druker, 2004; Pui et al., 2008).

About 25% of cases of B-cell precursor ALL, which is the most frequent form of acute leukaemia in children, harbor the *TEL-AML1* fusion gene generated by the t(12;21)(p13;q22) chromosomal translocation (Pui et al., 2004). Although the molecular pathogenesis of *TEL-AML1* positive leukaemia remains unclear, findings in mice establish the *TEL* gene as an important regulator of haemopoietic-cell development, essential for definitive hematopoiesis (Hock et al., 2004). Similarly, *AML1* gene is essential for definitive embryonic hematopoiesis (Okuda et al., 1996; Wang et al., 1996). Thus, the presence of the TEL-AML1 fusion protein in B-cell progenitors may lead to disordered early B-lineage lymphocyte development, a hallmark of leukaemic lymphoblasts (Pui et al., 2008).

1.2.5 Fusion genes in solid tumor

In contrast to hematologic malignancies where the existence and importance of fusion genes are well-documented, our knowledge of fusion genes is lacking in solid epithelial tumors. In solid tumor, efforts at fusion gene discovery are traditionally hampered by many challenges, ranging from technical difficulties (sample handling, preparation of cytogenetic spreads) to daunting levels of genomic complexity and clonal heterogeneity

in individual tumors that may confound standard cytogenetic assays (Heim and Mitelman, 2008).

Intriguingly, more than 80% of all known gene fusions are attributed to leukemias, lymphomas and bone and soft tissue sarcomas which account for only 10% of all human cancers. In contrast, common epithelial cancers, which account for 80% of cancer-related deaths, only attributed to 10% of known recurrent gene fusions (Kumar-Sinha et al., 2008; Mitelman et al., 2005). Gene fusions described among epithelial cancers have included *RET–NTRK1* fusions in papillary thyroid carcinoma, *PAX8–PPARG* in follicular thyroid carcinoma, *MECT1-MAML2* in mucoepidermoid carcinoma, *TFE3–TFEB* in kidney carcinomas, and *BRD4–NUT* in midline carcinomas (Kumar-Sinha et al., 2008). Remarkably, recurrent gene fusions have not been detected in the most prevalent carcinomas, such as prostate, breast, lung, and gastrointestinal tumours (Mitelman, 2000).

However, recently, one research group in Michigan University reported a recurrent fusion transcript *TMPRSS2-ERG* in prostate tumor and subsequently Soda *et al.* identified a fusion transcript *EML4-ALK* in non-small-cell lung cancer (Fig.1.7 and Fig.1.8) (Soda et al., 2007; Tomlins et al., 2005). The prevalence of gene fusions in these two types of cancers are ~ 50% and 5% respectively, suggesting that the prevalence of solid cancer subtypes characterized by recurrent gene fusions varies across different cancer types. Nevertheless, these studies established that oncogenic gene rearrangements do exist in common epithelial cancers and may be identified using high-resolution genomic approaches.

In 2010, we used transcriptome sequencing to identify *BRAF*-related gene fusions in GC, providing pioneering evidence for this important class of molecular aberrations in gastrointestinal cancer (Palanisamy et al., 2010).

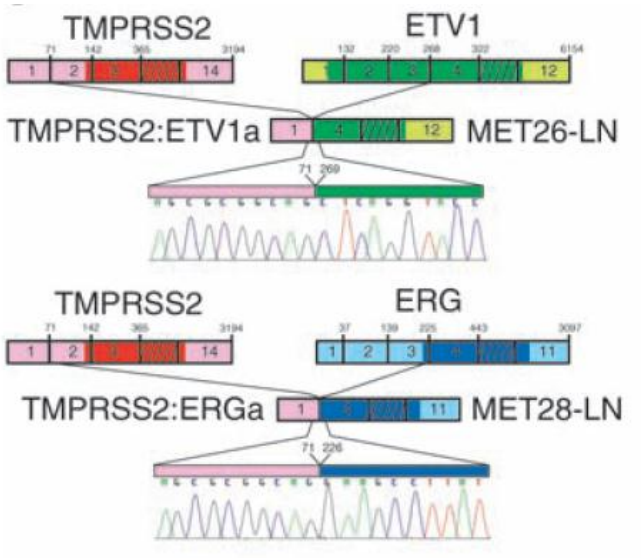


Figure 1.7. Identification of *TMPRSS2:ETV1* and *TMPRSS2:ERG* gene fusions in prostate cancer (PCA)
Schematic of *TMPRSS2* with *ETV1* fusion in MET26-LN and *ERG* fusion in MET28-LN.

Adapted from Science *310*, 644-648. (Tomlins et al., 2005)

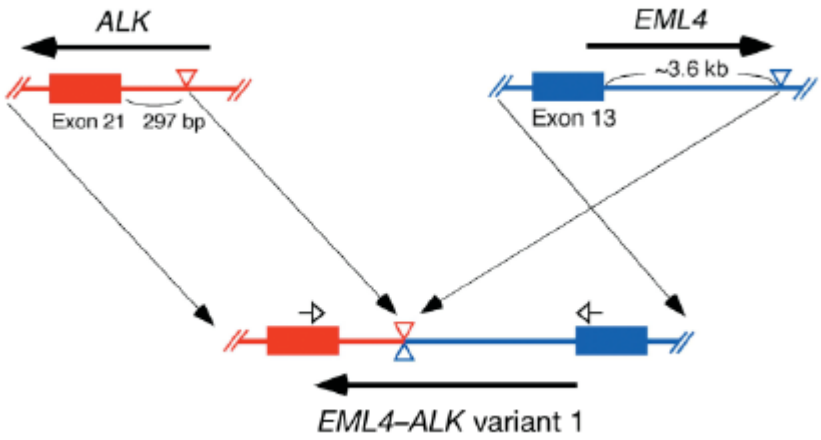


Figure 1.8. Gene fusion between *EML4* and *ALK* in Non-small-cell lung cancer
Both the *ALK* gene and the *EML4* gene map to chromosome 2p, but have opposite orientations.

Adapted from Nature *448*, 561-566. (Soda et al., 2007)

1.2.6 Fusion gene identification using genomic breakpoint analysis (GBA)

Identification of fusion genes has evolved with improvements in technology. One approach is genomic breakpoint analysis (GBA) using array based technology that allows the identifications of amplifications, deletions and allelic imbalances (Nannya et al., 2005; Tsukamoto et al., 2008). GBA identifies putative chromosomal breakpoints by examining closely spaced microarray probes displaying prominent transitions in copy number status, from low to high copy number or vice versa. We have discussed previously that gene fusions can arise from balanced or unbalanced chromosome rearrangements. However, genomic breakpoint analysis using high resolution array based technology only detects changes of gene dosage and is unable to identify balanced translocations (Fig.1.9).

Using genomic breakpoint analysis, one research group has recently identified novel *PAX5* fusion transcripts in leukemia using high-resolution SNP arrays, reiterating the ability of a high resolution genomic approach in identifying fusion genes (Fig.1.10) (Kawamata et al., 2008). Although such studies have demonstrated that GBA can discover novel fusion genes in leukemia, it is currently unclear if GBA can also be successfully applied to solid epithelial cancers, due to increased genomic complexity in such cancers.

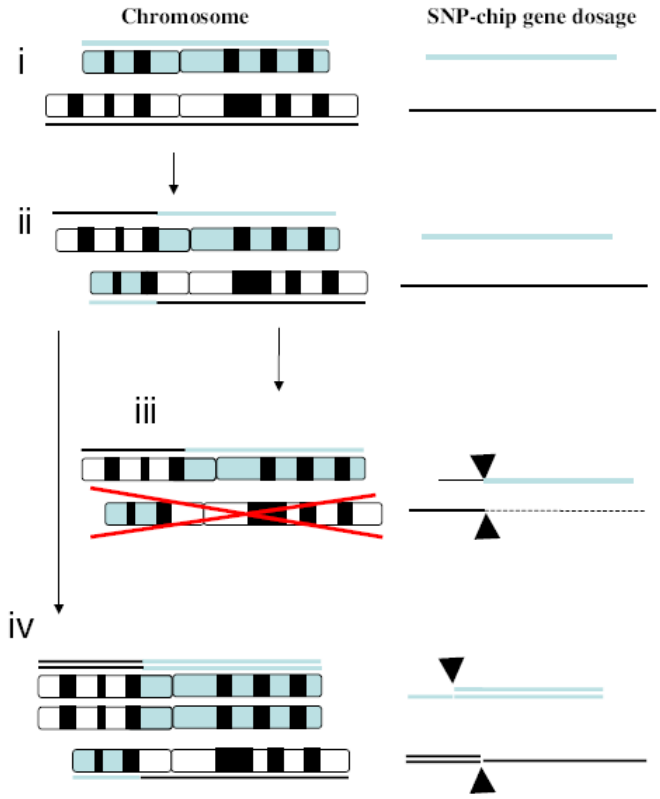


Figure 1.9. Copy number analysis detected genes involved in unbalanced translocations

Copy number analysis can identify breakpoints of translocations when one of the paired translocated chromosomes is either lost or duplicated/amplified. (*Left*) Chromosomal status. Gene dosages are indicated either above or beneath the chromosomes. (*Right*) Results of SNP-chip analysis.

i, normal chromosomes; gene dosage is normal.

ii, reciprocal translocation; gene dosage is normal.

iii, one of the paired translocated chromosomes is lost; gene dosage is lower than normal on the left side of the upper chromosome and the right side of the lower chromosome. Arrow heads indicate the breakpoint of the translocation in each chromosome.

iv, one of the paired translocated chromosomes is duplicated; gene dosage is higher than normal on the right side of the upper chromosome and the left side of the lower chromosome. Arrow heads indicate the breakpoint of this translocation in each chromosome.

Adapted from Proc Natl Acad Sci U S A *105*, 11921-11926. (Kawamata et al., 2008)

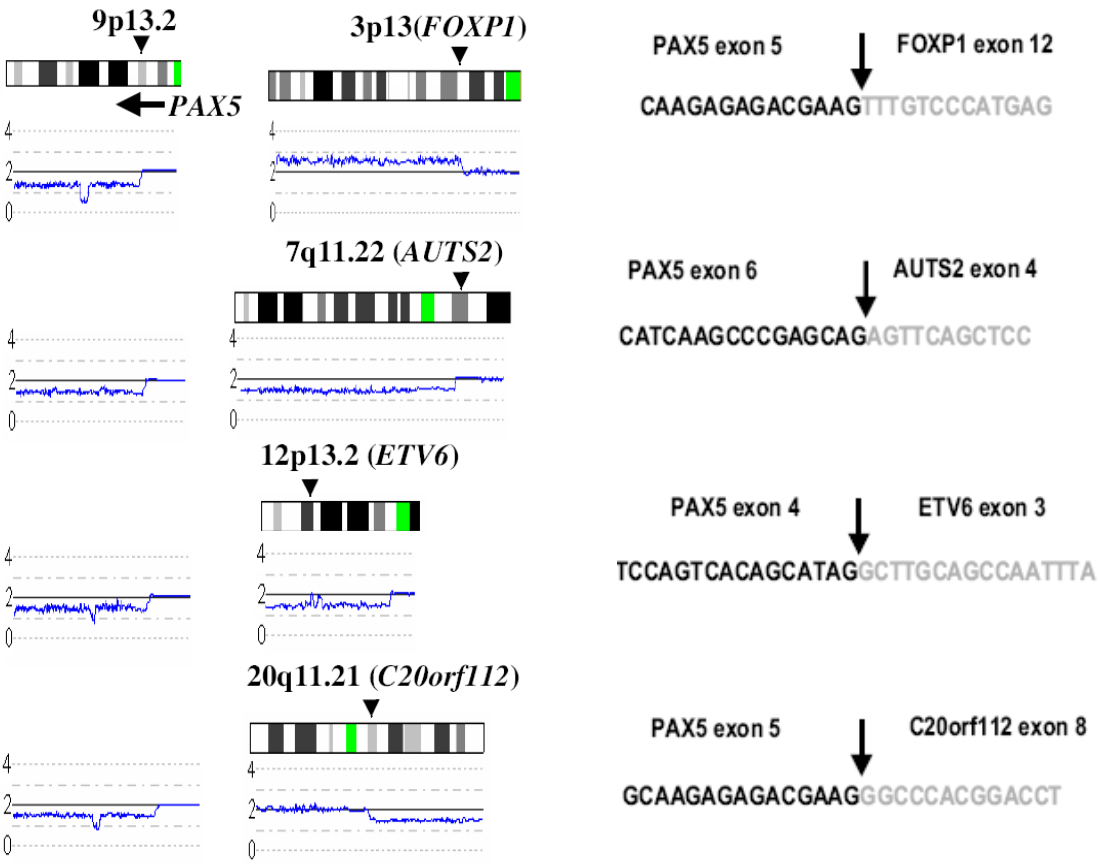


Figure 1.10. *PAX5* gene is fused to partner genes
Left, start sites of deletion at 9p13.2 involving the PAX5 gene. (*Left*) SNP-chip data of representative cases with 9p13.2 deletions. A vertical arrow indicates the start sites of 9p deletion that involves the *PAX5* gene. A horizontal arrow shows the direction of transcription of the *PAX5* gene. (*Right*) Chromosomal abnormalities of partner chromosomes. Arrow heads indicate the start sites of duplication or deletions. Genes involved in the start sites are shown.
Right, Fusion sequences of the *PAX5* and partner genes. Joining sequences of fused transcripts are shown from the indicated exon of the fused gene.

Adapted from Proc Natl Acad Sci U S A *105*, 11921-11926. (Kawamata et al., 2008)

1.3 CD44

1.3.1 CD44 family

CD44 is a transmembrane glycoprotein that acts as a hyaluronan-binding surface receptor. It is a major adhesion molecule for the extracellular matrix (ECM) and has been implicated in a wide variety of physiological processes (Ponta et al., 2003). Numerous isoforms of *CD44* exist through alternative mRNA splicing (Fig.1.11). They are encoded by a single, highly conserved gene located on chromosome 11, which contains 20 exons. 10 of which can be differentially spliced. All isoforms have 5 constant exons at the amino terminus followed by 10 alternatively spliced exons, 2 constant membrane-proximal coding exons, 1 constant transmembrane-coding exon, and 2 potentially alternatively spliced exons coding for cytoplasmic domains (Schmits et al., 1997). The great variability in splicing leads to at least 18 different *CD44* isoforms (Screaton et al., 1993; Tolg et al., 1993). The smallest *CD44* isoform, which is known as *CD44* standard (CD44s), is predominantly expressed in hematopoietic cells and normal epithelial cell subsets, whereas the larger variant isoforms (CD44v) with the membrane-proximal extracellular region are abundant in some epithelial-type carcinomas (Ponta et al., 2003; Tanabe et al., 1993).

1.3.2 Molecular function of CD44

CD44 may provide a link between the plasma membrane and the actin cytoskeleton. CD44 is among a group of proteins that crosslink transmembrane receptors to the actin cytoskeleton. The interplay between ERM (Ezrin, radixin, and moesin) proteins and

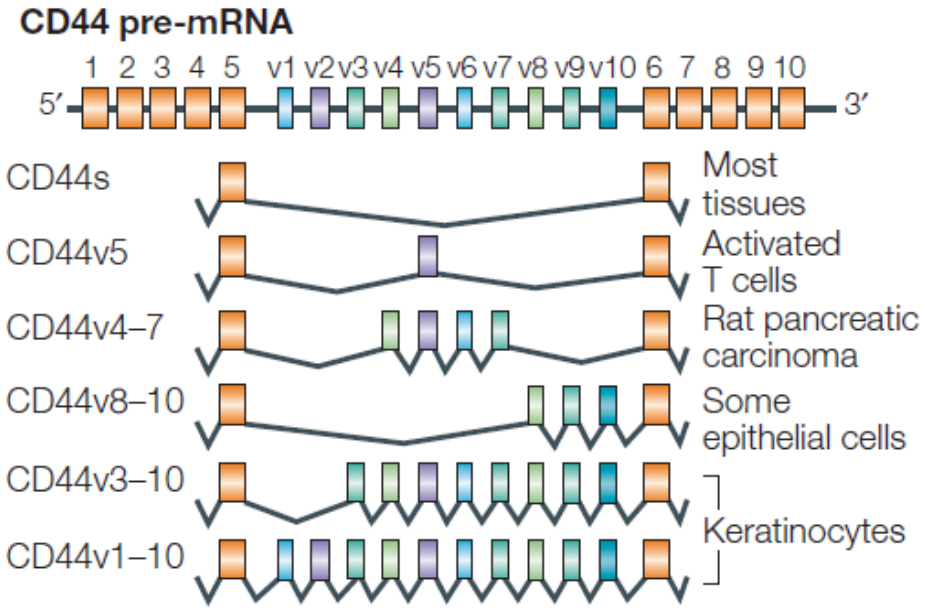


Figure 1.11. *CD44* transcripts
CD44 pre-mRNA is encoded by 20 exons, 10 of which can be regulated by alternative splicing (variant or 'v' exons). The smallest *CD44* isoform, which is known as *CD44* standard (CD44s), is ubiquitously expressed in vertebrates in developing and adult organisms, whreas the larger variant isoforms are expressed in some epithelial tissues and in several cancers.

Adapted from Nat Rev Mol Cell Biol *4*, 33-45. (Ponta et al., 2003)

CD44 might establish the dynamics that are required for cellular movement (Tsukita et al., 1994).

CD44 has co-receptor functions that mediate the signaling of receptor tyrosine kinases. This function of CD44 proteins is important for at least a small group of receptor tyrosine kinases, which includes Met and members of the ERBB family of receptor tyrosine kinases (RTK) (Orian-Rousseau et al., 2002).

CD44 binds ligands, which influences cellular behaviors independent of interactions with receptor tyrosine kinases or the actin cytoskeleton, these functions include its adhesion to hyaluronan and other components of ECM, as well as axon deterrence, its action as a platform for growth factors and other molecules, and its regulation of assembly, disassembly or uptake of pericellular hyaluronan-based matrices (Stamenkovic et al., 1991).

1.3.3 CD44 function in health and disease

CD44 is expressed in most human cell types and is implicated in a wide variety of physiological processes, including lymphocyte homing and activation, regulation of growth, survival, differentiation and motility (Nagano and Saya, 2004; Ponta et al., 2003). CD44 proteins have essential functions in life and altered expression or dysfunction causes pathogenic phenotypes.

In 1991, it was reported that an alternative spliced form of CD44 could confer metastatic capacity to a rat pancreatic carcinoma cell line. Since then, a vast number of publications have addressed the role of CD44 and CD44 splice variants in tumor progression and

metastasis and their relevance as diagnostic and prognostic parameters for human tumors, such as breast, colorectal and lymphomas (Combaret et al., 1996; Herrlich et al., 1993; Imazeki et al., 1996; Stauder et al., 1995). However, there are some tumor types including neuroblastomas and prostate carcinomas, in which the absence of CD44 variant expression correlates with transformation and poor prognosis, and overexpression of CD44 in prostate carcinoma cells even suppressed metastatic behavior (De Marzo et al., 1998; Gao et al., 1997; Shtivelman and Bishop, 1991). Although studies showed controversial results on CD44 function in cancers, substantial evidence suggested that aberrant expression of CD44 and CD44 variants are associated with many human tumors.

In addition, CD44 has recently been identified as one of the cell surface markers associated with cancer stem cells (CSCs) in several types of tumors (Al-Hajj et al., 2003; Collins et al., 2005; Dalerba et al., 2007). CSCs are malignant cell subsets in hierarchically organized tumors; they are selectively capable of tumor initiation and self-renewal and give rise to the bulk population of nontumorigenic cancer cells through differentiation. However, the underlying mechanisms for the emergence of CD44 positive CSCs during tumorigenesis have not been elucidated (Ishimoto et al., 2011).

1.3.4 CD44 and GC

Comparative genomic hybridization (CGH) study has detected frequent 11p13 amplifications in GC (Fukuda et al., 2000). The minimal common region at 11p13, within the 11p11.2–14 amplicon, harbors the *CD44* gene. In addition, expression of *CD44v* has been previously shown to be associated with progression of human gastrointestinal malignancies (Kim et al., 2009). Recently, *CD44v* is shown to be heterogeneously

expressed in mouse gastric tumors, being highly abundant in proliferative cells and slow-cycling stem-like cells (Ishimoto et al., 2010). CD44v may function to interact with a glutamate-cystine transporter, and control the intracellular level of reduced glutathione (GSH) in human gastrointestinal cancer cells. High levels of CD44 expression showed an enhanced capacity for GSH synthesis and defense against reactive oxygen species (ROS). Ablation of CD44 resulted in suppressed tumor growth in a transgenic mouse model of GC. These findings establish a function for CD44v in regulation of ROS defense and tumor growth in GC (Ishimoto et al., 2011).

1.4 Excitatory amino acid transporters (EAATs)

1.4.1 Glutamate and EAATs

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system (CNS). It is estimated that ~ 40% of all neurons distributed throughout the CNS involve this amino acid (Bunch et al., 2009). Glutamate plays a central role in normal brain functions including cognition, memory, and learning (Riedel et al., 2003). However, excessive elevation of the extracellular glutamate concentration mediates excitotoxicity and causes neuronal cell death (Beal, 1992). Glutamate has been implicated in many pathological conditions such as epilepsy, cerebral ischemia, myotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease and schizophrenia (Amara and Fontana, 2002; Danbolt, 2001). There is no enzymatic system available for metabolizing glutamate in the extracellular space; the only way to maintain glutamate homeostasis is through glutamate uptake via glutamate transporters (Danbolt, 2001; Wu et al., 2010).

Excitatory amino acid transporters (EAATs) are high-affinity, sodium-dependent glutamate carriers from solute carrier family 1 (SLC1) in the CNS, which maintain extracellular glutamate concentrations and contribute to the clearance of glutamate released during neurotransmission (Gebhardt et al., 2010). At least five sodium-dependent glutamate transporter subtypes have been identified and characterized: EAAT1 (SLC1A3) (Storck et al., 1992), EAAT2 (SLC1A2) (Pines et al., 1992), EAAT3 (SLC1A1) (Kanai and Hediger, 1992), EAAT4 (SLC1A6) (Fairman et al., 1995), and EAAT5 (SLC1A7) (Arriza et al., 1997). EAAT3 and EAAT4 are localized predominantly in neurons, and EAAT5 is enriched in retinal tissue, whereas EAAT1 and EAAT2 are generally expressed abundantly on astrocytic plasma membranes associated with

excitatory synaptic contacts (Sims and Robinson, 1999). Among the five known human EAAT subtypes, EAAT1 and EAAT2 have the greatest impact on clearance of glutamate released during neurotransmission. Studies of EAAT3, EAAT4 and EAAT5 suggest more subtle roles for these subtypes in regulating excitability and signaling (Amara and Fontana, 2002).

1.4.2 EAAT2/SLC1A2

EAAT2 (Hereafter refer as *SLC1A2*) gene has been mapped to chromosome 11p13(Li and Francke, 1995), showing a genomic structure of 11 protein coding exons, spanning more than 55 kb genomic DNA (Meyer et al., 1997). Several homologous transcript variants of *SLC1A2* have been identified, showing a heterogeneity of the 5' untranslated sequence and a variability of the N-terminal amino acids of the putative proteins (Arriza et al., 1994; Meyer et al., 1998; Munch et al., 1998). The *SLC1A2* cDNA contains an open reading frame of 1722 bp encoding for a predicted protein of 574 amino acids (Meyer et al., 1998).

SLC1A2 is responsible for more than 90% of glutamate uptake in the brain (Boudker et al., 2007; Gebhardt et al., 2010). It is predominantly expressed in astrocytes and functions to transport glutamate into astrocytes for conversion into glutamine, which is then released and recycled by neurons to generate glutamate. This process is thought to contribute to energy metabolism in the brain, because if glutamate was not recycled by conversion into glutamine and had to be synthesized repeatedly, glucose consumption in the brain would be much higher (Lauriat and McInnes, 2007; Marcaggi and Attwell, 2004). Dysregulations of SLC1A2 has been associated with multiple psychiatric and

neurological disorders such as ischemia, Alzheimer's disease and Huntington's disease (Su et al., 2003; Ye et al., 2010). These findings emphasize the importance of glutamate transport and the SLC1A2 transporter in astrocytes to normal brain function and their association with multiple pathologic changes in the brain (Su et al., 2003).

1.4.3 Glutamate and cancer metabolism

The Warburg effect, describes the propensity for cancer cells and tissues to take up glucose avidly and convert it almost exclusively to lactate (aerobic glycolysis). This effect has been proposed to have a central role in cancer cell metabolism (Kroemer and Pouyssegur, 2008; Vander Heiden et al., 2009). Although the glycolytic pathway generates ATP and produces metabolic intermediates for cancer cells, glucose can only provide carbon source. Glutamine is another essential nutrient for cancer cells and is an abundant amino acid in the serum. Essential functions of glutamine include its conversion to glutamate as a metabolic intermediate to be channeled into the tricarboxylic acid (TCA) cycle and its function as a precursor for the biosynthesis of nucleic acids, certain amino acids, and glutathione (Lu et al., 2010) (Fig. 1.12).

There is an expanding body of literature suggesting that glutamate is likely to behave as a critical amino acid metabolite in many cancers (Rzeski et al., 2001; Takano et al., 2001). Glutamate levels have been shown to be elevated in many cancers including GC, and glutamate deprivation has been shown to sensitize cancer cells to apoptotic stimuli (Okada et al., 1993; Rothstein et al., 1996). The exact mechanistic requirement of cancer cells for glutamate is an active area of study, and may be related to the Warburg effect. It has been proposed that the Warburg effect is a result of cancer cells having a deficiency

in their ability to shuttle glycolytic metabolites into the Kreb's cycle to generate ATP. Because glutamate (and its sister amino acid glutamine) can be converted intracellularly into alpha-ketoglutarate, the availability of glutamate could thus provide a secondary route to facilitate ATP production via the Kreb's cycle (Yang et al., 2009).

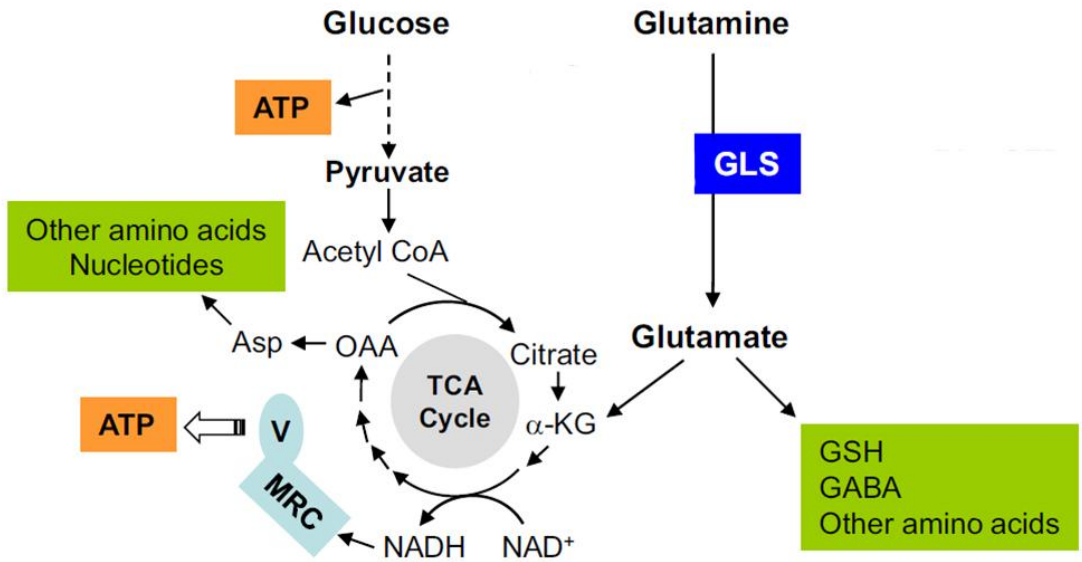


Figure 1.12. Glutamine metabolism
The interconnection between glutamine metabolism and glucose metabolism is also shown. GLS, glutaminase; TCA cycle, tricarboxylic acid cycle; MRC, mitochondrial respiratory chain; V, mitochondrial respiratory complex V; OAA, oxaloacetate; Asp, aspartate; a-KG, a-ketoglutarate.

Modified from Cancer Cell *18*, 199-200. (Lu et al., 2010)

1.5 Rationale of the study

Fusion genes are of exceptional interest to the cancer biology community due to their cancer specific nature (being produced by genomic aberration events). They represent ideal drug targets and diagnostic markers (Mitelman et al., 2007). In blood cancers, fusion genes are routinely used to diagnose particular clinical subtypes, and treatment of CML, once a uniformly lethal cancer, has been revolutionized by gleevec, which targets the *BCR-ABL* fusion gene (Druker et al., 2006; Hughes et al., 2003; O'Brien et al., 2003). In contrast to the blood cancers, few recurrent fusion genes have been identified in solid epithelial cancers (eg gastric, colon, lung and breast) to date. However, recent studies such as *TMPPRS2-ERG* (prostate cancer) and *EML4-ALK* (non-small-cell lung cancer) have shown that fusion genes do exist in solid tumors, and that these entities may be uncovered using high-resolution genomic approaches (Soda et al., 2007; Tomlins et al., 2005). Thus, we intend to apply high resolution genomic approaches to identify possible fusion transcripts in GC, the second highest cause of global cancer mortality (Hartgrink et al., 2009).

The specific aims for the current study are listed below:

1. Identify genes located at the breakpoints of chromosomal aberrations in human gastric cancer cell lines (GCCLs) and gastric tumors using GBA
2. Identify possible fusion transcripts using cell line models
3. Perform functional studies on recurrent fusion transcripts in *vitro* and in *vivo*
4. Examine recurrence of the fusion in primary gastric tumors
5. Investigate the clinical impact of the fusion to GC

Chapter II: Material and Methods

2.1 Primary tissues and cell lines

Primary gastric tumors and normal tissues were obtained from the Singhealth Tissue Repository, an institutional resource of National Cancer Centre of Singapore and Singapore General Hospital. All patient samples were obtained with informed patient consent and approvals from Institutional Review Boards and Ethics Committees. Gastric cancer cell lines (GCCLs) AGS, KATO III, SNU1, SNU16, N87, and Hs746T were purchased from the American Type Culture Collection (ATCC). AZ521, Ist1, TMK1, MKN1, MKN7, MKN28, MKN45, MKN74, Fu97, and IM95 cells were obtained from the Japan Health Science Research Resource Bank. SCH cells were provided by Yoshiaki Ito (Cancer Sciences Institute of Singapore). YCC cells were a gift from Sun-Young Rha (Yonsei Cancer Center, South Korea). HFE145 cells were a gift from Hassan Ashktorab (Howard University College of Medicine, United States).

2.2 Cell culture

2.2.1 Culture of gastric cancer and normal cell lines

SNU16 cells were maintained in RPMI 1640 (National Cancer Center Medium Prep, Singapore) containing 10% heat-inactivated Fetal Bovine Serum (FBS) (Gibco, Invitrogen Singapore), 1% Penicillin-Streptomycin (Gibco, Invitrogen Singapore), 0.1 mM Non-Essential Amino Acids Solution (Gibco, Invitrogen Singapore) and 2 mM L-glutamine (Gibco, Invitrogen Singapore). Cells were cultured in 75 cm^2 tissue culture flasks (Corning Life Sciences, USA) in 37^oC incubator with 5% carbon dioxide in a water saturated environment. Cell morphology was monitored every day to ensure that healthy cells were used for the experiments. When cells reach 90% confluency for subculture, cells were collected and centrifuged for 3 min at 1,500 rpm. Then supernatant was removed and the cell pellet was washed with warm sterile Phosphate-Buffered saline (PBS) (National Cancer Center Medium Prep, Singapore) and centrifuged for 3 min at 1,500 rpm. After removing supernatant, cell pellet was resuspended in medium and divided into new tissue culture wares with a ratio of 1:4.

AGS cells were maintained in the same medium and condition as SNU16 cells mentioned above. Cell morphology was monitored every day to ensure that healthy cells were used for the experiments. When cells reach 90% confluency for subculture, the medium was discarded and the adherent cell monolayer was gently washed with PBS. Appropriate amounts of 0.25% trypsin-EDTA **(**National Cancer Center Medium Prep, Singapore**)** was added to the cell monolayer until most of the cells became detached. Subsequently,

supernatant was removed and the cell pellet was resuspended in medium and divided into new tissue culture wares with a ratio of 1:4.

HFE145 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (National Cancer Center Medium Prep, Singapore) containing 10% heat-inactivated FBS, 1% Penicillin-Streptomycin, 0.1 mM Non-Essential Amino Acids Solution and 2 mM L-glutamine. Cells were maintained in T75 tissue culture flasks in 37^{o}C humidified incubator with 5% carbon dioxide through subcultures as a ratio of 1:4. Subculturing of cells was performed as AGS cells described above.

2.2.2 Quantification of cell number

Cell viability was determined by mixing 10 µl of diluted cell suspension with 10 µl of Trypan blue solution (Sigma Aldrich, USA). Approximately 10 µl of the cell mixture was loaded into C-Chip disposable hemocytometer (Digital Bio Technology, UK). Viable cells were counted under CKX41 microscope (Olympus, USA) and corresponding cell number was calculated accordingly.

Alternatively, 10 µl of the cell mixture (cell suspension and trypan blue solution) was loaded into the Invitrogen Countess^{TM} counting chamber (Invitrogen, Singapore). The chamber was then analysed in the Invitrogen Countess^{TM} automated cell counter (Invitrogen, Singapore). The value indicating cell concentration of the mixture was shown and multiple 2X of the value to determine the actual concentration of the original cell suspension.

2.3 DNA isolation

DNAs were isolated from primary gastric tumor and normal tissues and gastric cancer cell lines using Qiagen Blood and Cell Culture DNA extraction kit (QIAGEN, Germany). DNA yields were determined using Nanodrop 2000 (Thermo Scientific Inc. USA) and DNA quality was determined using a 0.5% agarose gel.

2.3.1 DNA extraction from primary gastric tissues

Primary tissue samples were cut and weighed accordingly. Tissues were grind to fine powder with liquid nitrogen in a pre-cooled mortar and pestle. Tissues were homogenized with 2 ml Buffer G2 (with 4 µl RNase A). 0.1 ml proteinase K stock solution were added to the homogenate and mixed by vortexing. Samples were incubated at 50°C overnight.

After incubation, the samples were used for genomic DNA extraction using Genomic-tip Protocol following manufacturer's instruction. Briefly, QIAGEN Genomic-tip 20/G was equilibrated with 1 ml Buffer QBT. After equilibration, samples were applied to the genomic-tip and entered the resin by gravity flow. Next, genomic-tip was washed with 1 ml Buffer QC for three times. The genomic DNAs were eluted with 2 x 1 ml of Buffer QF. 1.4 ml isopropanol was added to the eluted DNA to precipitate the DNA. Samples were mixed and centrifuged immediately at >5000 *g* for at least 15 min at 4°C. Supernatant was carefully removed. Subsequently, DNA pellets were washed with 1 ml, cold 70% ethanol and centrifuged again at >5000 *g* for 10 min at 4°C. Supernatants were removed without disturbing the pellets. Pellets were air-dried for 5–10 min and resuspended in 0.1–2 ml of TE (pH 8.0) buffer. DNA was dissolved overnight on a shaker.

2.3.2 DNA extraction from cultured cells

Frozen cell pellets were washed with PBS and resuspended in 0.5 ml PBS. 0.5 ml ice cold buffer C1 and 1.5 ml of distilled water were added to the cell suspension. The samples were mixed by inverting the tube several times and incubated for 10 min on ice. After incubation, lysed cells were centrifuged at 4°C for 15 min at 1300 *g*. Supernatants were discarded. Subsequently, 0.25 ml ice-cold Buffer C1 and 0.75 ml ice-cold distilled water were added to the pellets. The pelleted nuclei were resuspended by vortexing and centrifuged again at 4°C for 15 min at 1300 *g*. 1 ml Buffer G2 was added to the pellets and completely resuspended by vortexing for 10–30 sec at maximum speed. After vortexing, 25 µl Proteinase K stock solution was added and the samples were incubated at 50°C water bath for 30–60 min.

After incubation, the samples were used for genomic DNA extraction using Genomic-tip Protocol (described above).

2.4 Agilent 244k aCGH profiling and Genomic Breakpoint Analysis (GBA)

106 primary tumors and 27 cell lines were profiled using Agilent 244K Human Genome Microarrays (Agilent Technologies, Santa Clara, CA), which contained *In situ* synthesized 60 mer oligo probes covering both the coding and non-coding sequences of total genomic DNA with an average spatial resolution of approximately 6 kb. Briefly, Genomic DNA isolated from samples and control genomic DNAs (human spleen DNA) were labeled with Cy3-dUTP and Cy5-dUTP (Agilent Genomic DNA Labeling Kit PLUS) respectively (Agilent Technologies, Santa Clara, CA). Subsequently, entire reaction product was hybridized onto the microarray using the Agilent aCGH Hybridization Kit (Agilent Technologies, Santa Clara, CA). Microarrays were washed using the Agilent aCGH Wash Kit (Agilent Technologies, Santa Clara, CA). For each DNA probe on the microarray, the ratio of intensity of the fluorescence measured for the two fluors was determined by scanning through Agilent DNA Microarray Scanner (Agilent Technologies, Santa Clara, CA). A red color represents increased DNA copy number, green represents decreased copy number (i.e., deletion), and yellow represents no change in DNA copy number in tumor cell DNA compared with normal cell DNA. Data were extracted using Agilent's Feature Extraction version 9.1 software and analyzed using Agilent's CGH Analytics version 3.5 software (Agilent Technologies, Santa Clara, CA). Genomic breakpoint analysis was then carried out using a *Z*-score algorithm with a threshold of 2.0 and a 1 pt window to identify putative genomic breakpoints.

2.5 Fluorescence *in-situ* Hybridization (FISH)

SNU16 interphase and metaphase cell pellets were prepared for FISH analysis by standard hypotonic treatment and fixation after colcemide exposure (10 µg/ml) for 2 hrs. Prior to hybridization, cells were pre-treated with pepsin (100 mg/ml) (Sigma, USA) and 0.01 mol/L HCl at 37°C (5 min), fixed in 1% formaldehyde (Sigma, USA) (10 min), and dehydrated in an ethanol series. Fosmid and BAC probes were obtained from BACPAC Resource Center (BPRC, CHORI, Oakland, California, USA), and grown following vendor instructions. DNA was extracted with Nucleobond PC500 (Macherey-Nagel), followed by labeling with either biotin-16-dUTP (Roche, Germany) or digoxigenin-11-dUTP (Roche, Germany) using an Enzo Nick Translation DNA labeling system. Approximately 20 ng of each probe was used per hybridization in addition to 10 ug of Cot1-DNA (Invitrogen, USA). The slides and probes mixes were codenatured on a hot plate set at 75°C and hybridized overnight at 37°C. Post hybridization washes were performed at 45°C in pre-warmed formamide 50%/2X SSC solution (twice), followed by two washes in 2X SSC (twice). Slides were blocked with blocking reagent (Roche, Germany), followed by incubation with avidin-conjugated fluorescein isothiocyanate (FITC) (Roche, Germany) and anti-Digoxygenin-Rhodamine (Roche, Germany) respectively. DAPI counterstain (Vector Laboratories, USA) was then used to stain the nuclei to enable visualization. Slides were mounted with vectashield (Vector Laboratories, USA). Fluorescence images were captured with a 60X objective using a cooled charge-coupled device (CCD) camera attached to a Nikon fluorescence microcscope. Automated images capture was performed using ISIS software (Metasystems, Germany).

2.6 RNA isolation

Total RNA were extracted from primary gastric tissues and gastric cancer cell lines using Qiagen RNeasy® mini kit (QIAGEN, Germany). RNAs were quantitated using either a Nanodrop 2000 (Thermo scientific Inc. USA) or Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

2.6.1 RNA extraction from primary gastric tissues

Primary tissue samples were cut and weighed accordingly. Tissues were grind to fine powder with liquid nitrogen in a pre-cooled mortar and pestle. Tissue powder was lysed with 600 µl RLT buffer (with β-mercaptoethanol). The samples were spinned for 2 min at full speed. The lysed supernatant was then transferred directly into a QIAshredder spin column placed in a 2 ml collection tube and spinned for 2 min at full speed at room temperature. 600 µl of 70% ethanol were added to the homogenized lysate and mixed well. Subsequently, samples were transferred to RNeasy spin column and centrifuged for 15 sec at 8000 *g*. After discarding the flow-through, 700 µl Buffer RW1 was added to the RNeasy spin column, and centrifuged for 15 sec at 8000 *g* to wash the spin column membrane. Subsequently, 500 µl buffer RPE was added to the column and centrifuged for 2 min at 8000 *g* (twice). 30-50 µl RNase free water was then added to the column to elute RNA.

2.6.2 RNA extraction from cultured cells

Briefly, frozen cell pellets were lysed with 600 µl RLT buffer (with β-mercaptoethanol). The lysates were transferred directly into a QIAshredder spin column placed in a 2 ml

collection tube and spinned for 2 min at full speed at room temperature. 600 µl of 70% ethanol were added to the homogenized lysate and mixed well. Subsequently, samples were transferred to RNeasy spin column and centrifuged for 15 sec at 8000 *g*. After discarding flow-through, 700 µl Buffer RW1 was added to the RNeasy spin column, and centrifuged for 15 sec at 8000 *g* to wash the spin column membrane. Subsequently, 500 µl buffer RPE were added to the column and centrifuged for 2 min at 8000 *g* (twice). 30-50 µl RNase-free water was then added to the column to elute RNA.

2.7 RLM-RACE (RNA-Ligase Mediated Rapid Amplification of cDNA Ends)

RLM-RACE was performed using the FirstChoice® RLM-RACE kit (Applied Biosystems, USA) following manufacturer's instruction.

2.7.1 5' RACE

10 µg of total RNA was first treated with 2 µl Calf Intestinal Alkaline Phosphatase (CIP) and incubated at 37°C for one hour to remove 5' phosphate groups, followed by Tobacco Acid Pyrophosphatase to remove 5' cap structures. After RNA linker ligation, mRNA transcripts were reverse transcribed using M-MLV reverse transcriptase (Applied Biosystems, USA). To amplify first strand cDNAs, outer 5' PCR was performed using 5' RACE outer primers (provided in kit) and a *SLC1A2* exon 3 primer (ACACACTGCTCCCAGGATGA) with SuperTaq™ Plus polymerase (Applied Biosystems, USA). Subsequently, inner 5' PCR was performed using a 5' RACE inner primer (provided in kit) and a *SLC1A2* exon 2 primer (AGCCAAGATGACTGTCGTGCATTC). Typical PCR reaction cycle consists of the following steps: denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 30 sec, annealing at the required temperature for 30 sec or its variable depending on the product size and 72°C for 1-2 min and final extension at 72°C for 7 min. After gel electrophoresis, PCR bands of interest were excised and cloned into pCR® 2.1-TOPO® (Invitrogen, USA) vectors (see below for details). Purified plasmid DNAs were sequenced bi-directionally on an ABI 3730 automated sequencer (Applied Biosystems, USA). A minimum of 5 independent colonies were sequenced in each experiment.

2.7.2 3' RACE

Briefly, 1 μg of total RNA was reverse transcribed using a 3'RACE adaptor and reverse transcriptase provided in the kit. To amplify first strand cDNAs, outer 3' PCR was performed using 3' RACE outer primers and a *SLC1A2* exon 1 primer (TTGAGGCGCTAAAGGGCTTACC) with SuperTaq™ Plus polymerase (Applied Biosystems, USA). Subsequently, inner 3' PCR was performed using a 3' RACE inner primer (provided in kit) and a separate *SLC1A2* exon 1 primer (CAGACCATGGCATCTACGGAAGG). After gel electrophoresis, PCR bands of interest were excised and cloned into PCR® 2.1-TOPO® (Invitrogen, USA) vectors. Purified plasmid DNAs were sequenced bi-directionally on an ABI 3730 automated sequencer (Applied Biosystems, USA). A minimum of 5 independent colonies were sequenced in each experiment.

2.8 Semi-quantitative reverse-transcription PCR (RT-PCR)

GC RNAs were reverse transcribed by SuperScript II reverse transcriptase enzyme using oligo-dT (T18) primers (Invitrogen, USA). To detect *CD44-SLC1A2*, RT-PCR was performed using forward primers to *CD44* exon 1 (CCATGGACAAGTTTTGGTGGCA) and reverse primers to either *SLC1A2* exon 3 (GTATATCCCCTGGGAAGGCT); exon4 (CAGCTGCTTCTTGAGCTTGGGA); exon 5 (AAGCAGGCTTGGACAAGGTT) or exon 6 (CTCGTTCAACAGAGAGACAACAGC). Products were resolved by gel electrophoresis and bands of interest were excised and cloned for subsequent analysis. To evaluate wild-type *CD44* and *SLC1A2* expression independently of *CD44-SLC1A2*, we used *CD44* forward primer targeting exons 3-5 (Sense: AGTCACAGACCTGCCCAATGC; antisense: TGCTGTCTCAGTTGCTGTAGCA); *SLC1A2* primers targeting exon 1 (Sense: ATCGCCTGCAAATCCCCAGC; antisense: TGCCACCTGTGCTTTGCTGC). GAPDH was used as equal loading control (Sense: GCTCTCCAGAACATCATCCCTGC; antisense: TCTAGACGGCAGGTCAGGTCCAC). Standard PCR was performed in a 50 µl reaction using Bio-rad Tetrad 2 thermal cycler (Biorad, USA). Each reaction included 25 µl GoTaq® Hot Start Colorless Master Mix (2X Colorless GoTaq® Reaction Buffer (pH 8.5), 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dTTP and 4 mM $MgCl_2$) (Promega, USA), 2 µl of 10 µM forward primer, 2 µl of 10 µM antisense primer, 2 µl template DNA and 19 µl of RNase free water. PCR reaction cycle consists of the following steps: denaturation at 94°C for 3min, followed by 35 cycles of 94°C for 30 secs, annealing at the required temperature for 30 sec or its variable depending on the product size and 72°C for 1-2 min and final extension at 72°C for 7 min.

2.9 Gel purification

Extraction and purification of DNA was performed using QIAquick gel extraction kit (QIAGEN, Germany) following the manufacturer's protocol. Briefly, the DNA fragment was excised from the agarose gel with a scalpel and weighed accordingly. 3 volumes of Buffer QG were added to 1 volume of gel (100 mg ~ 100 µl) and incubated at 50°C for 10 min to completely dissolve gel. After the gel slice has dissolved completely, 1 gel volume of isopropanol was added to the sample. Next, the sample mixture was transferred directly to a QIAquick spin column placed in a 2 ml collection tube and centrifuged for 1 min at maximum speed. The flow-through was discarded and 0.75 ml of buffer PE was added to wash the column and centrifuged again for 1 min (twice). Finally, 50 µl of RNase free water was added to the column to elute purified DNA.

2.10 DNA cloning techniques for sequencing

Gel purified DNA products were cloned into pCR® 2.1-TOPO® (Invitrogen, USA) vectors for downstream sequencing analysis using TOPO TA Cloning® for sequencing kit ((Invitrogen, USA) following manufacturer's protocol.

2.10.1 DNA ligation

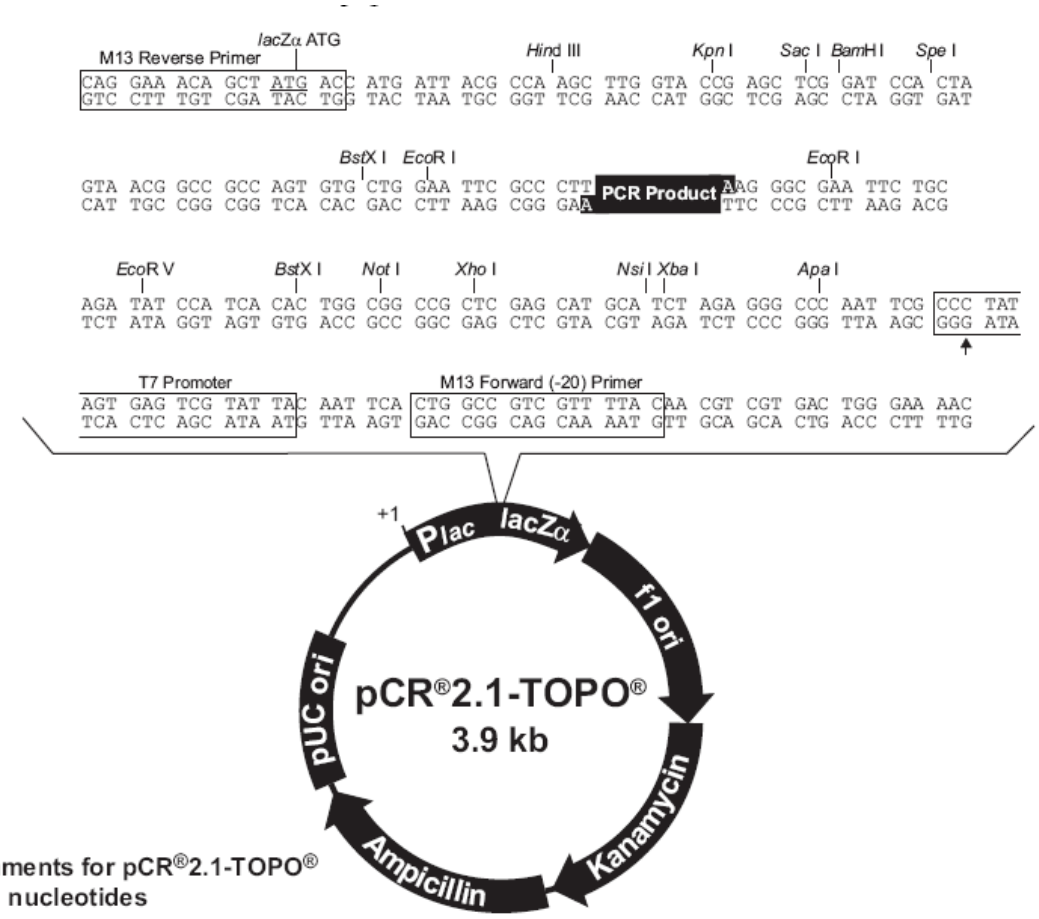
DNA TOPO ligation reaction was carried out in 6 µl of volume, containing 4 µl PCR product, 1 µl salt solution and 1 µl PCR® 2.1-TOPO® vectors (Invitrogen, USA). The ligation reaction was incubated at room temperature for 5 min.

2.10.2 Transformation

2 µl of the TOPO ligation reaction was added to one vial of One Shot Mach1™-T1R Competent *E.coli* Cells (Invitrogen, USA) and incubated on ice for 30 min. The mixture was heated at 30 sec at 42^{o}C and cooled immediately on ice for 2 min. 250 µl of S.O.C. medium (Invitrogen, USA) was added to the mixture and incubated with shaking (200rpm) at 37^{o}C for 1 hour. After incubation, ¼ of the transformation reaction mixture was spread onto LB plate supplemented with 50 µg/ml kanamycin (Invitrogen, USA) and the plate was incubated at 37^{o}C overnight (200rpm).

2.10.3 Plasmid purification

Plasmid DNA isolation was carried out using PureLink™ Quick Plasmid Miniprep Kit (Invitrogen, USA) for subsequent sequencing analysis. Briefly, One Shot Mach1™-T1R Competent *E.coli* Cells containing desired plasmid were cultured in antibiotics-supplemented LB broth, shaked and incubated at 37^{o}C for 4 hour (200rpm). After incubation, 1-3 ml bacterial culture was transferred to a 1.5 ml microcentrifuge tube and centrifuged for 15 min at 1500 *g*. 240 µl of resuspension solution was added to completely resuspend the pellets, 240 µl lysis buffer was added and incubated for 4 min at room temperature. Subsequently, 340 µl neutralization buffer was added and mixed gently. The mixture was immediately centrifuged for 10 mins at max speed to clarify the cell lysates. The supernatant was transferred to a spin column and centrifuged at room temperature at 10,000-14,000 *g* for 1 min. 650 µl of wash buffer was then added to wash the column. Finally, 50 µl of water was added to the center of the column to elute DNA. DNA yield and purity was determined using Nanodrop.



Comments for pCR®2.1-TOPO®
3931 nucleotides

*Lac*Zα fragment: bases 1-547
M13 reverse priming site: bases 205-221
Multiple cloning site: bases 234-357
T7 promoter/priming site: bases 364-383
M13 Forward (-20) priming site: bases 391-406
f1 origin: bases 548-985
Kanamycin resistance ORF: bases 1319-2113
Ampicillin resistance ORF: bases 2131-2991
pUC origin: bases 3136-3809

Fig. 2.1 pCR®2.1-TOPO® vector map. (Invitrogen, USA)

2.11 DNA sequencing

If direct PCR products were used for sequencing, Exonuclease I-Shrimp Alkaline Phosphatase (Exo-SAP) (Affymetrix, USA) was first added to the PCR samples to remove residual single-stranded primers and dNTPs. The mixture were incubated at 37°C for 1 hour and 72 °C for 15 min to inactive the enzyme activity.

Exo-SAP treated PCR products or purified DNA were then used for cycle sequencing reaction using ABI PRISM^{TM} BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems Inc. USA). The 10 µl reaction mixture containing 2 µl of 5X sequencing buffer, 0.32 µl of 10uM forward primer, 0.32 µl of reverse primer, 2-4 µl DNA template, 1 µl Bigdye terminator 3.1 and RNase free water.

The program for sequencing reaction was 10 sec at 96°C, 5 sec at 50°C, 4 min at 60°C for 30 cycles on the thermocycler. After the completion of cycles, the mixtures were purified using the ethanol precipitation method. Briefly, 4 µl of EDTA (62.5 nM) (Gibco, USA) and 30 µl pre-chilled pure ethanol was added to each sample. The solution was centrifuged for 40 min at 4000rpm at 4 ^{o}C and the supernatant was removed completely. The pellet was then washed by adding 90 µl of 70% pre-chilled ethanol. The samples were centrifuged for 20 min at 4000 rpm at 4 ^{o}C before being dried. Prior to sequencing, 10 µl of Hi-Di Formamide (Applied Biosystems, USA) was added to the DNA pellet and the mix was then used for sequencing on ABI PRISM 3730 Genetic Analyzer (Applied Biosystems, CA). Chromatograms were analyzed by SeqScape V2.5 and manual review.

2.12 Fiber-FISH

SNU16 cells and control cells (normal lymphoblastoid CCL159) were grown in RPMI 1640 enriched with 15% FBS, 1% PS and 1% L-Glutamine. 2-3ml of each cell suspension was centrifuged at 1200 rpm for 12 mins and then washed with 6 ml PBS twice. Pellets were diluted with PBS to a final concentration around 2-3 x 10^4/ml. 10 µl of each cell suspension was spread on a poly-L-lysine (Sigma, USA) coated slide, air dried and then fitted into a Cadenza Coverslip according to manufacturer's recommendations (Thermo Shandon). 150 µl of freshly made lysis solution (5:2 70 mM NaOH:absolute Ethanol) was applied to the slides, followed by 150 µl of 96% Ethanol. Slides were air dried at room temperature, treated with 3:1 Acetic Acid:Ethanol Fixative for 5 mins and dehydrated in Ethanol series (70%, 90%, and 100%) for 3 mins each. The FISH procedure was then applied.

2.13 Long range genomic PCR

CD44/SLC1A2 chromosomal inversions were detected using Qiagen LongRange PCR kit (QIAGEN, Germany) following the manufacturer's instructions. Briefly, 50 µl of reaction mixture was composed with 5 µl of 10X LongRange PCR buffer, 2.5 µl dNTP, 2 µl of forward primer at *CD44* exon 1 (GAAGAAAGCCAGTGCGTCTC, positive strand), 2 µl of reverse primer at *SLC1A2* intron 1 in the minimal breakpoint region (GAGGGCTGTCCTTAACGCCTAGC, negative strand), 0.4 µl LongRange PCR Enzyme Mix, 100 ng DNA template and RNase-free water.

PCR conditions for 0.1–10 kb sample were: initial activation at 93°C for 3 min, followed by 35 cycles of 93°C for 15 sec, annealing at the required temperature for 30 sec,

extension 1 min/kb 68°C (Use an extension time of 1 min per kilobase DNA for genomic DNA) and end of cycle run at 4°C.

PCR conditions for >10kb sample were: initial activation at 93°C for 3 min, followed by first 10 cycles of 93°C for 15 sec, annealing at the required temperature for 30 sec, extension for 1 min/kb 68°C (Use an extension time of 1 min per kilobase DNA for genomic DNA) and next 28 cycles of 93°C for 15 sec, annealing at the required temperature for 30 sec, extension for 1 min/kb +20 sec in each additional cycle at 68°C.

2.14 Quantitative-RT PCR (qRT-PCR)

SNU16 GC cells lines and 9 primary gastric tumors were selected for qRT-PCR analysis using Quantifast SYBR green PCR kit (QIAGEN, Germany) following manufacturer's instructions. T1, T2, T3 were group 1 gastric tumors which are 11p13 amplification negative and fusion negative; T4, T5, T6 are group 2 tumors which are 11p13 amplified but do not express *CD44-SLC1A2*; T7, T8, T9 were group 3 tumors expressing *CD44-SLC1A2* but are non-11p13 amplified. Briefly, 2 µg RNA was reverse transcribed by SuperScript III reverse transcriptase enzyme using oligo-dT (T18) primers (Invitrogen, USA). 25 µl of q-PCR reaction was prepared with 12.5 µl 2X QuantiFast SYBR Green PCR Master Mix, 1 µl of forward primer, 1 µl reverse primer, 2 µl template cDNA and RNase-free water. Primers used were: Fusion forward primer targeting *CD44* exon 1 (TTCGGTCCGCCATCCTCGTC) and reverse primer targeting *SLC1A2* exon 2 (CACTTCCACCTGCTTGGGCA); *SLC1A2* exon 1 forward primer (GCCCGTTGAGGCGCTAAAGG) and reverse primer (AGCACTATCCGGCAGCTGTG); and *GAPDH* (internal control) forward primer

(CCACCCAGAAGACTGTGGATGG) and reverse primer (CACTGACACGTTGGCAGTGG). qPCR reaction was carried out using cycling program outlined below: initial activation at 95ºC for 5 min, 35 cycles of denaturation for 10 sec at 95ºC and Combined annealing/ extension for 30 sec at 60ºC. Samples were analyzed using Applied Biosystem 7900HT system (Applied Biosystems, USA).

To avoid nonspecific binding of SYBR Green to any double strand DNA or primer dimmers, primers were carefully designed to give a product around 200 base pairs. Experimental reports including amplification analysis, melting curve analysis and threshold cycle number were provided automatically.

2.15 Protein isolation

Total proteins, cytoplasm proteins and membrane phase proteins were extracted from gastric cancer cells to serve different experimental purposes.

2.15.1 Total protein isolation from cell lysates

Briefly, cells (1-5x10^6 /sample) were harvested by removing cell culture medium. The cell pellets were rinsed with 1X PBS and centrifuged to remove the supernatant. 1X RIPA buffer containing phosphatase inhibotors (Tris-HCl 1 M pH 7.5, 10% NP-40, 10% Na-deoxycholate, 500 mM NaCl, 0.5 M EDTA) were added to lyse the cells for 1 hour at 4ºC. The supernatant was transferred to a clean tube after centrifugation at 13,000 *g* at 4ºC and stored at -80ºC until used. Protein concentration was determined before storage.

2.15.2 Membrane phase extraction from cell lysates

Cells (1-5x10^6 /sample) pellets were resuspended with lysis buffer (10 mM Tris.Cl pH 7.5, 150 mM NaCl, 1%Triton X-114) at 4ºC for 1 hour before centrifuge at 800 *g* for 15 min at 4ºC. Supernatants were transferred to a new tube and incubated at 30ºC for 5-10 min until condensation occurred. The samples were then centrifuged at 300 *g* at room temperature for 5 min. The integral membrane proteins were in the detergent phase at the bottom of the centrifuge tube, while the supernatant containing the cytoplasm proteins that are not associated with Triton X-114.

The two phases were then collected separately and necessary amount of either Triton X-114 or base buffer (without Triton X-114) were added to obtain equal volume and approximately the same chemical concentration for both samples.

2.15.3 Determination of protein concentration

Protein concentration was measured using DC™ Protein Assay kit (BIO-RAD, USA). Briefly, 1 ml sample mixture was prepared (Including 100 µl reagent A, 100 µl water, 800 µl reagent B). Four dilutions of a protein standard were prepared using BSA. Protein samples were added to the mixture and incubated at RT for 15 min in dark. Protein concentration was measured by Lowry methods using Pharmacia Biotec Ultrospec 3000 UV/Visible spectrophotometer (GE healthcare, USA).

2.16 Western blotting

Western blotting was used for desired protein detection. Protein samples were prepared by the methods described in Protein Isolation chapter.

2.16.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE gel was carried out using a Mini-PROTEIN Electrophoresis System (BIO-RAD, USA). The gels were prepared dependent on different concentrations in order to detect proteins of various molecular weights. Briefly, separating gels were prepared by mixing appropriate amount of 30% acrylamide/bisacrylamid solution (37.5:1, BIO-RAD, USA) with water, 1.5 M Tris Cl (pH 8.8), 10% SDS, 10% fresh ammonium persulfate (APS) (BIO-RAD, USA) and TEMED (BIO-RAD, USA). Propanol was covered on top of the separating gel. After polymerization, overlay was decanted and a 4% stacking gel was poured. 2 ml of 4% stacking gel was prepared by mixing 1.35 ml water, 0.33 ml 30% acrylamide/bisacrylamid solution, 0.5 ml Tris.Cl (pH 6.8), 100 µl 10% SDS, 100 µl APS and 4 µl TEMED. Comb was inserted correctly and allowed to polymerize completely before running.

Unless stated otherwise, 60 µg of protein for each sample was resolved on SDS-PAGE gels under denaturing conditions. The protein samples were mixed with 3X loading buffer (150 mM Tris HCl pH 6.8, 6% SDS, 0.3% bromophenol blue, 30% glycerol and 300 mM mercap (add fresh)) and boiled in 100ºC heat blocker for 5 min. samples were then loaded and the gel was run under 80 volts till the dye reached the resolving gel and thereafter at 120 V till satisfactory protein separation was observed. Kaleidoscope Protein Plus protein standards (BIO-RAD, USA) was run as a protein weight reference.

2.16.2 Gel transfer

The separated protein was electrophorectically transferred onto pre-activated PVDF membranes (BIO-RAD, USA). The polyacrylamide gel was rinsed in transfer buffer (25 mM Tris,192 mM Glycine, 0.1% SDS, 20% methanol), placed in-between layers of filter paper and membrane of the same size of the gel, inserted into the plastic holder with sponge soaked in transfer buffer. The air bubbles were removed by rolling with a glass pipet. The transfer was performed in transfer buffer at 100 V for 1 hour at 4ºC.

2.16.3 Immunoprobing and detection

The membranes were incubated with anti- SLC1A2/EAAT2 primary antibody (1:500, Cell Signaling Technology, Danvers, MA) for two hours at room temperature, followed by anti-rabbit HRP-conjugated secondary antibody (Cell Signaling Technology, MA) for another 2 hours. The membranes were washed extensively in PBST washing buffer (1X PBS and 0.1% Tween-20). For equal loading of samples, the blots were probed with anti-α-tubulin primary antibody (Cell Signaling Technology, MA) and then anti-rabbit HRP-conjugated secondary antibody (Cell Signaling Technology, MA). The membrane was then washed before incubating with the Pierce ECL Western Blotting Detection System (Thermo Scientific, USA) by mixing equal volumes of the two reagents for 5 min and exposed to Hyperfilm X-ray films (Kodak, USA) at various desired time lengths in the dark room.

2.17 Immunofluorescence staining

Cells were fixed with 3.7% formaldehyde (prepared in 1xPBS) and permeabilised with 0.1% Triton-X 100 (prepared in 1xPBS), then deposited on microscopy slides in a cytospin centrifuge. After 3 times washes with 1X PBS, cells were blocked with 1% BSA. Subsequently, Fluorescent labeling was performed using the anti-SLC1A2 antibody (Cell Signaling Technology, MA) as primary for 2 hour and FITC-conjugated as secondary antibody (Sigma-Aldrich, Singapore) for 2 hour. Nikon Eclipse TE2000-U microscope was used to analyze the fluorescent microscopy staining, and images were captured using a Leica DC 300F camera (Nikon, Japan).

2.18 siRNA transfection

GC cells were stably transfected with either specific siRNAs targeted to the *CD44-SLC1A2* fusion site (siRNA1: CGCAGAUCGUGCCAACAAUUU; siRNA2: GCACAUCGUGCC AACAAUAUU) (100 nM, custom siRNA siGENOME with SMART selection, Dharmacon) or negative control scrambled siRNAs using siPORT^{TM} *NeoFX*^{TM} transfection reagent (Applied Biosystem, USA) in Opti-MEM Medium (Invitrogen, USA) following the manufacturer's protocol. Briefly, appropriate number of cells was resuspended and cultured in 2.3 ml normal growth medium in a 6 well plate prior to transfection. 5 µl siPORT^{TM} *NeoFX*^{TM} transfection reagent was mixed well with 100 µl of Opti-MEM medium and incubated at RT for 10 mins. Meantime, 7.5 µl of 10 µM siRNA was also mixed with 100 µl of Opti-MEM medium. After 10 min incubation, the two solutions were mixed together and incubated for 10 min at RT to allow transfection complexes to form. Subsequently, the mixture was added to the cells and the

cells were incubated at 37ºC for 48 hour. Stable transfectants were selected with puromycin for 4 weeks.

For wild-type *SLC1A2* siRNAs transfection, GC cells were transfected with specific non-overlapping siRNAs targeted either to *SLC1A2* exon 1 or *SLC1A2* downstream regions (100 nM, custom siRNA siGENOME with SMART selection, Dharmacon) or negative control scrambled siRNAs using siPORT^{TM} *NeoFX*^{TM} transfection reagent (Applied biosystem) in Optimem Medium (Invitrogen, Singapore) as described above.

2.19 *CD44-SLC1A2* DNA cloning and overexpression

The full length coding regions of *CD44–SLC1A2* cDNA were inserted into the pEGFP-N1 vector (Clontech, USA) and transformed into One Shot Mach1^{TM}-T1R Competent *E.coli* Cells as described above. Plasmid DNA was purified and control vectors or fusion-GFP vectors were introduced into gastric cells using SuperFect® Transfection Reagent (QIAGEN, Germany). Briefly, the day before transfection, 2–8 x 10^5 cells were seeded in 2 ml appropriate growth medium in 6-well plates and incubated under normal growth conditions. On the day of transfection, 2 µg DNA were diluted with 100 µl Opti-MEM medium. 20 µl SuperFect Transfection Reagent was then added to the DNA solution and incubated for 5-10 min at RT to allow transfection-complex formation. While complex formation took place, cells were washed with PBS and 500 µl normal cell growth medium were added to the reaction tube containing the transfection complexes. The complexes were carefully mixed and immediately transferred to the cells. Cells were incubated with the transfection complexes for 2–3 h under their normal growth conditions before changing the medium to fresh cell culture medium. Cells were then incubated for 48 hours. Stable transfectants were selected using g418 for 4 weeks.

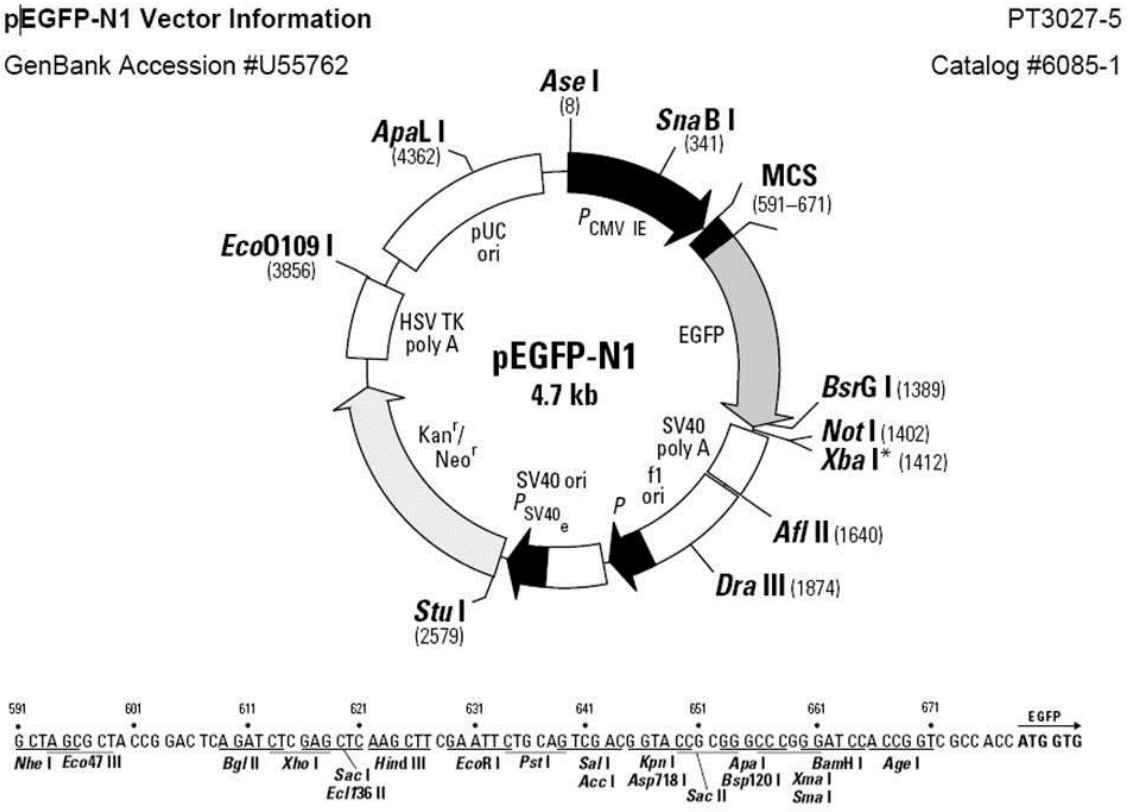


Fig. 2.2 Restriction map and multiple cloning site of pEGFP-N1 vector. (Clontech)

2.20 *In vitro* cell assays

2.20.1 Cell proliferation assay

Cell proliferation assays were performed using a CellTiter96® Aqueous Nonradioactive Cell Proliferation Assay kit (MTS) (Promega, USA) following the manufacturer's instructions. Briefly, gastric cells (1000-2000 cells per well) were cultured overnight in 96-well plate in 100 µl of normal growth medium at 37ºC. 20 µl of MTS reagents were added into each well and mixed well with the culture medium. The mixture was incubated for 3 hours at 37ºC in dark. After 3 hours incubation, proliferation rates were measured at 490 nm using a PerkinElmer plate reader (PerkinElmer, USA). Each assay was performed in triplicate, and the results were averaged over three independent experiments.

2.20.2 Cell invasion assays

Cell invasion assays were performed using Biocoat^{TM} Matrigel^{TM} invasion chambers with 8 µm pore filter inserts (BD Bioscience, USA). Briefly, 0.5 ml of normal growth medium were added to both the interior of the inserts and bottom of wells for rehydration for 2 hours. After rehydration, the medium was carefully removed without disturbing the layer of Matrigel™ Matrix on the membrane. Subsequently, 0.75 ml of FBS was added as chemoattractant to each well of the Companion Plate. The chambers and control inserts were carefully transferred to the wells with chemoattractant without causing air bubbles. Immediately, $2.5x10^4$ cells in 0.5 ml normal growth medium were added to the 24-well chambers. The invasion chambers were then incubated for 22 hours in a humidified tissue

culture incubator, at 37^{o}C, 5% CO_2 atmosphere. After incubation, the invading cells were counted using light microscopy.

2.20.3 Soft agar assays

Base layers constituted of 0.5% Gum Agar in 1x McCoy's 5A and 10% FBS were poured into 6-well plates and allowed to set. Meantime, appropriate number of cells was resuspended in normal growth medium. 10 µl of cell suspension was mixed well with 1.5 ml top agar constituted of 0.3% Gum Agar in 1x McCoy's 5A and 10% FBS and plated to each well immediately. The agar was allowed to set for at least 30 min before 2 ml of normal growth medium was seeded on top of the solidified base layer. The plates were incubated at 37°C in humidified incubator for 3-4 weeks, during which plates were fed drop-wise with complete media. After 3-4 weeks, plates were photographed using the Kodak GL 200 System (EpiWhite illumination) (Kodak, USA). Each assay was performed in triplicate, and the results were averaged over three independent experiments.

2.20.4 Glutamate assays

Gastric cancer cells and primary tissues were lysed in glutamate assay buffer and glutamate concentrations were determined using a Glutamate Assay Kit (BioVision, USA). Briefly, cells or primary tissues were homogenized in 100 µl assay buffer centrifuged to remove insoluble material at 13,000 g for 10 minutes. Subsequently, 10 ul of each serum samples were added directly to each well with duplicates. Glutamate standard were prepared in duplicate to generate 0, 2, 4, 6, 8, 10 nmol/well standard. A vendor-provided glutamate Enzyme Mix was added to each well which recognizes

glutamate as a specific substrate leading to proportional color development. The plate was then measured at 450 nm in a Tecan infinite M200 microplate reader (Tecan, USA).

2.20.5 Drug treatments

For cisplatin treatments, appropriate number of cells was seeded into 96 well plates after siRNA transfection. Subsequently, cisplatin or 5-FU at increasing dosages were added (0-1 mM) to respective wells. Cells were subjected to MTS proliferation assays (described above) after 48 hours of drug treatment. Each assay was performed in triplicate, and the results were averaged over three independent experiments.

2.21 Copy number analysis (Affymetrix)

Affymetrix SNP6 arrays were processed using Affymetrix GTC 4.0 software and tumor profiles were normalized against a matched normal reference. The data was visualized using Nexus 5.0 software (Biodiscovery, USA). A Rank Segmentation algorithm, a variation of the segmentation method based on Circular Binary Segmentation (S*1*) was used to segment the copy number data across the genome.

2.22 Gene expression analysis

Gene expression data is available from the Gene Expression Omnibus under access number GSE15460. Gene expression profiles (Affymetrix U133P2 arrays) were normalized using the MAS5 algorithm. Comparisons between *CD44* and *SLC1A2* expression values was performed on a subset of 45 samples for which gene expression, copy number information, and *CD44-SLC1A2* gene fusion status was available. Unsupervised clustering was based on all probesets after removing the bottom 25% of probes with the lowest interquantile range. Hierarchical clustering and Wilcoxson signed rank tests were performed using R software 2.9.0. FDR q value calculations were calculated using the R package 'qvalue'. GO analysis was performed using the DAVID database (S*2, S3*).

2.23 Statistical analysis

Experiments were assessed by Student's unpaired *t* test, with the exception of the tumor/normal glutamate measurements where a paired t-test was used. *P* values <0.05 were considered statistically significant.

Chapter III: RESULTS

3.1 Analysis of GC copy number alterations identifies recurrent *SLC1A2/EAAT2* genomic breakpoints

Previous cytogenetic and chromosomal studies have established that individual GCs often exhibit high degrees of karyotypic complexity (Espinoza et al., 1999; Tay et al., 2003). Motivated by the intimate link between copy number alterations (CNAs), chromosomal rearrangements and gene fusions (Bignell et al., 2007; Persson et al., 2009), we hypothesized that a detailed fine-scale survey of genomic CNAs might reveal potential genes disrupted by fusion events in GC.

3.1.1 Validation of Agilent 244k aCGH data

Using high density microarrays (Agilent Human 244k), we profiled a discovery cohort of 133 GCs (106 primary tumors and 27 cell lines). Validation of the microarray data was achieved by comparing the CNA profiles to earlier studies. We successfully re-identified many previously known genomic aberrations in GC, including amplifications in *c-Myc*, *HER2*, *RAB23*, and deletions in *PTEN* (Fig. 3.1) (Hou et al., 2008; Mitsui et al., 2007; Sato et al., 2002; Varis et al., 2002).

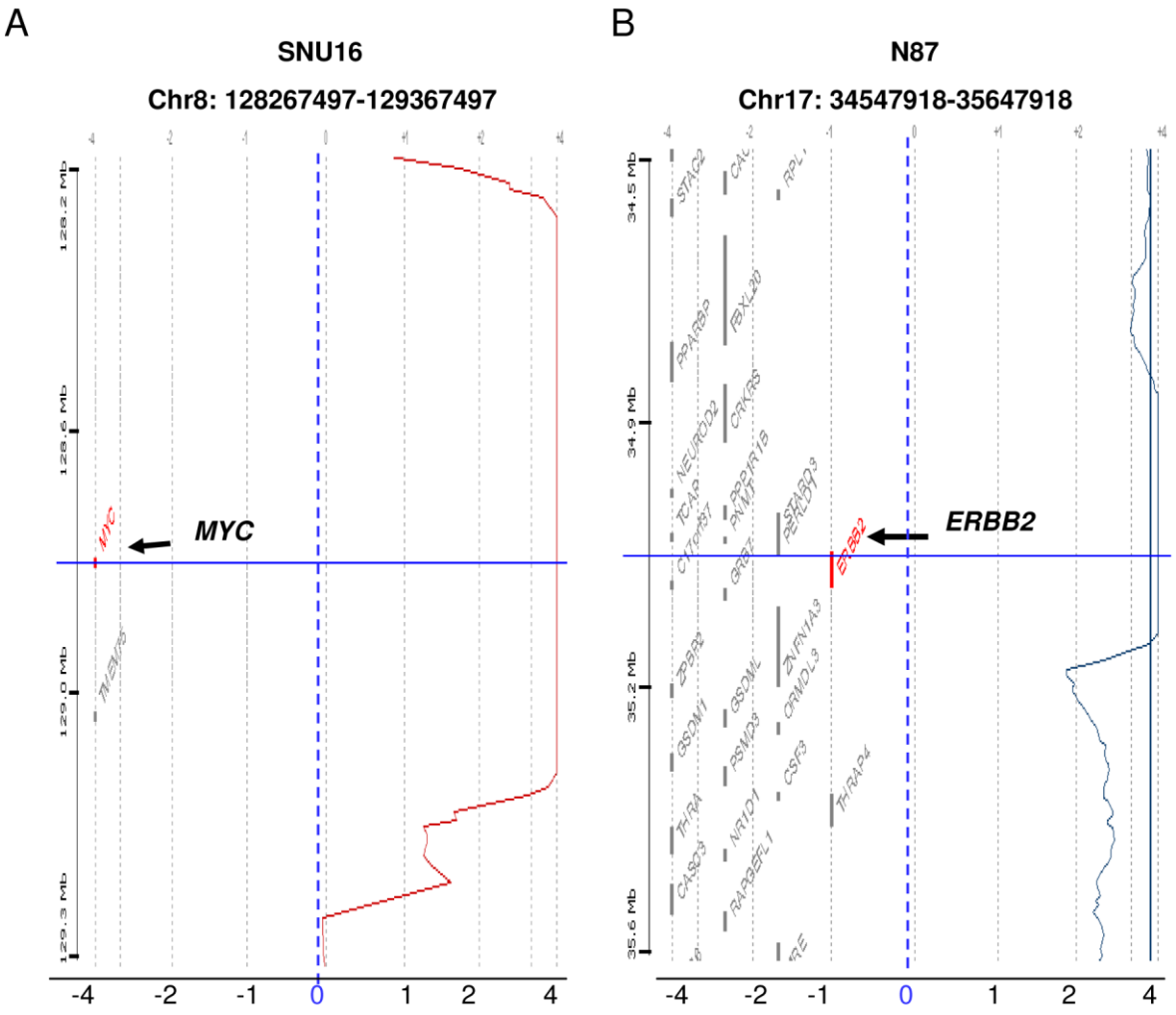


Figure 3.1. *MYC*, *ERBB2*, *RAB23*, and *PTEN* genomic aberrations in SNU16, N87, HS746T and TMK1 cells detected by Agilent 244k aCGH

A. Copy number amplification of the *MYC* oncogene detected in SNU16 cells. X-axis, log2-transformed smoothened copy number values averaged over a 0.1Mb moving window. Y-axis, physical genomic coordinates along chromosome. The black arrows point to the respective genes. Vertical Blue dashed lines indicate copy number values of 2 (Log-transformed copy number = 0).

B. *ERBB2* amplification in N87 GC cells. X-axis, log2-transformed smoothened copy number values averaged over a 0.1Mb moving window. Y-axis, physical genomic coordinates along chromosome. The black arrows point to the respective genes. Vertical Blue dashed lines indicate copy number values of 2 (Log-transformed copy number = 0).

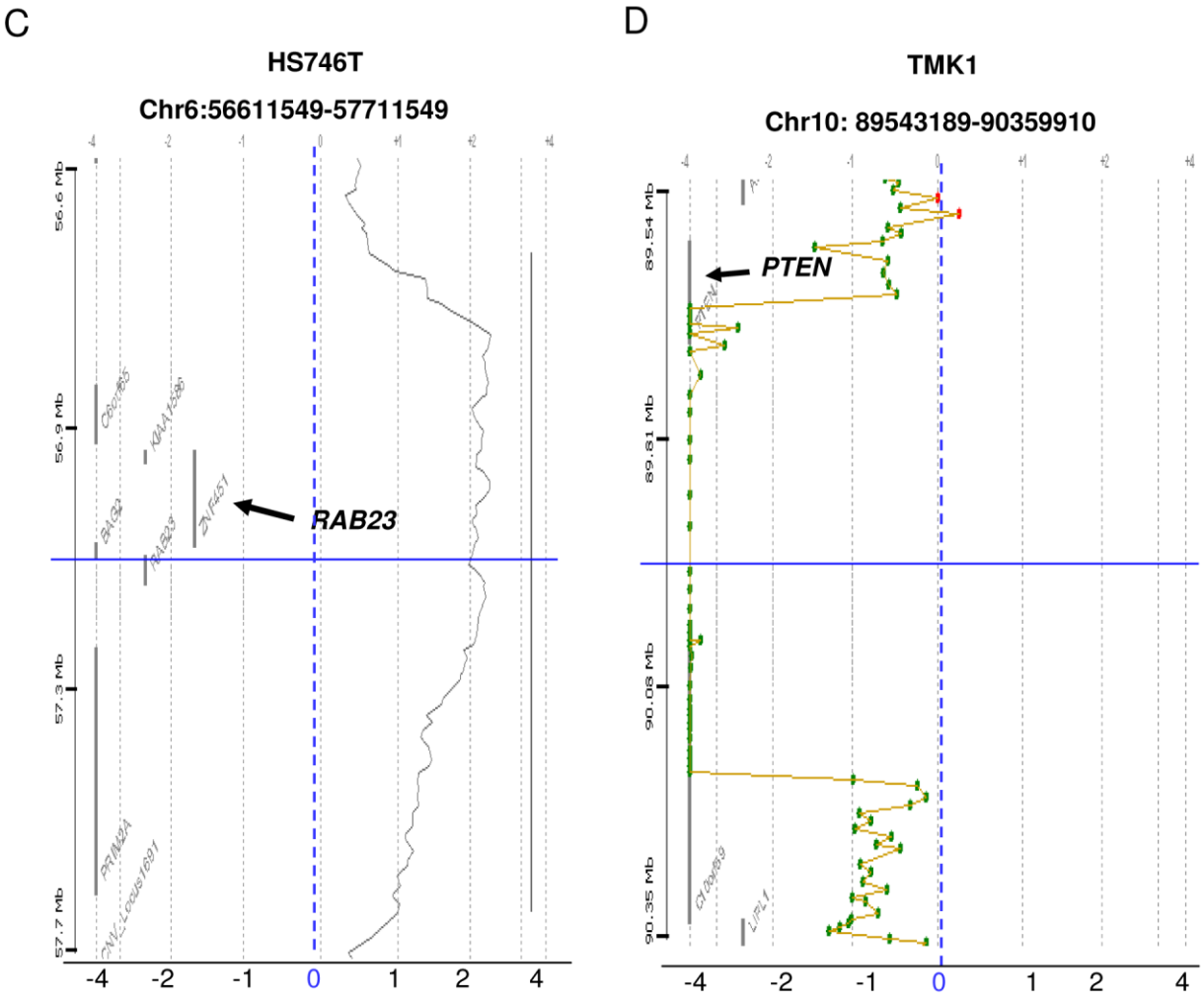


Figure 3.1. *MYC, ERBB2, RAB23*, and *PTEN* genomic aberrations in SNU16, N87, HS746T and TMK1 cells detected by Agilent 244k aCGH

C. *RAB23* amplification in HS746T GC cells. X-axis, log2-transformed smoothened copy number values averaged over a 0.1Mb moving window. Y-axis, physical genomic coordinates along chromosome. The black arrows point to the respective genes. Vertical Blue dashed lines indicate copy number values of 2 (Log-transformed copy number = 0).

D. *PTEN* deletion in TMK1 cells. X-axis, log2-transformed smoothened copy number values. Y-axis, physical genomic coordinates along chromosome. The black arrows point to the respective genes. Vertical Blue dashed lines indicate copy number values of 2 (Log-transformed copy number = 0).

3.1.2 Breakpoint analysis using aCGH data reveals recurrent *SLC1A2/EAAT2* genomic breakpoints

To nominate potential fusion genes, we employed a technique called Genomic Breakpoint Analysis (GBA), previously used to identify fusion genes in leukemia (Kawamata et al., 2008). In this strategy, putative chromosomal breakpoints were identified by examining closely-spaced microarray probes displaying prominent transitions in copy number status, from low to high copy number or vice versa. In total, we identified 90 genomic breakpoints occurring in genes such as *CALCR*, *PERLD1*, and *CKAP5* (intragenic breakpoints, Table 3.1). The highest recurrence rate among the breakpoints was 3% and most breakpoints occur only in one sample. This is probably due to the high genomic complexity and clonal heterogeneity of GC. A representative example of a genomic breakpoint, occurring in the *CALCR* gene, is shown in Figure 3.2A.

For the majority of genes exhibiting genomic breakpoints in multiple samples (eg *CRKRS*, *TTC25*), the breakpoints were randomly scattered throughout the gene body consistent with a random breakage model of chromosomal amplification. However, four GCs out of 133 (three primary tumors and one cell line - GC980417, GC20021048, GC2000038 and SNU16) exhibited genomic breakpoints specifically localized to the 5' region of the *SLC1A2/EAAT2* gene, encoding a high affinity glutamate transporter (hence after referred to as *SLC1A2*) (Fig. 3.2 B).

Table 3.1: List of genes exhibiting genomic breakpoints

Gene	Location	Genomic Event	Tumor/ Cell Line ID
CRKRS	17q12	Focal Amplification	90929219,2000484,73499299 , 2001140
SLC1A2	11p13	Focal Amplification	980417, 20021048, 2000038, SNU16
ZNFN1A3	17q21	Focal Amplification	20020164,9874831, 54115380, YCC9, N87
GSDML	17q12	Focal Amplification	90929219, 54115380
WIRE	17q21.1-17q21.2	Focal Amplification	2000484, 2001140
SMARCE1	17q21.3	Focal Amplification	2000484, 2001140
CKAP5	11p11.2	Focal Amplification	2000763, 2000563
PERLD1	17q12	Focal Amplification	20020164, 54115380
LDLRAD3	11p13	Focal Amplification	20021048, 20020448
TACC2	10q26	Focal Amplification	980417, 20020700
KIAA0319	14q13.2	Focal Amplification	YCC9, N87
TTC25	17q21.2	Focal Amplification	YCC9, N87
CDH26	20q13.33	Deletions	MKN74,MKN28
TRERF1	6p21.1-p12.1	Focal Amplification	YCC1, YCC6
PEX7	6q23.3	Amplification	990020
KIF 6	6p21.2	Amplification	990428
PDSS2	6q21	Focal Amplification	980447
POPDC3	6q21	Focal Amplification	980447
FLJ20294	11p11.12	Focal Amplification	980447

ELF5	11p13-p12	Focal Amplification	2000038
RASSF8	12p12.3	Focal Amplification	990428
DYM	18q12-21.1	Focal Amplification	990221
LOC126248	19q13.11	Focal Amplification	980255
CAT	11p13	Focal Amplification	20021048
c13orf24	13q22.1	Amplification	980021
BCAT1	12p12.1	Focal Amplification	980024
PDE1A	2q32.1	Amplification	980024
LST-3TM12	12p12.2	Focal Amplification	980024
IDH1	2q33.3	Amplification	20020164
ZNF403	17q12	Focal Amplification	20020164
SOCS7	17q12	Focal Amplification	20020164
RAB3IP	12q14.3	Amplification	20020563
RFFL	17q12	Focal Amplification	20020700
KLHL10	12q14.3	Focal Amplification	20020700
CALCR	17q12	Focal Amplification	20020700
MULK	17q21.2	Amplification	2001098
DSCR3	21q22.2	Focal Amplification	2000710
TP53I11	11p11.12	Focal Amplification	2000763
FLJ46154	11p14.1	Amplification	2000763
KIF14	1q32.1	Amplification	90929219
TMEM116	12q24.13	Focal Amplification	90929219
MAPKAPK 5	12q24.12-q24.13	Focal Amplification	90929219

C17ORF63	17q11.2	Amplification	9874831
TNKS	8p23.1	Focal Amplification	42761681
FBXL20,	17q12	Focal Amplification	54115380
PPARBP	17q12-q21.1	Focal Amplification	54115380
TSHZ2	20q13.2	Deletions	73499299
SLC24A3	20p13	Amplification	73499299
SCARA5	8p21.1	Amplification	20020448
RC74	8p21.1	Amplification	20020448
MTSS1	8p22	Focal Amplification	20020448
PROK2	3p13	Deletions	37175329
ST7	7q31.1-q31.3	Focal Amplification	20020070
WNT2	7q31.2	Focal Amplification	20020070
STARD3	17q11-q12	Focal Amplification	2001140
CCR7	17q12-21.2	Focal Amplification	2001140
PREP	6q16.3-q22.1	Focal Amplification	980447
RPA2	1p35.1	Amplification	YCC9
RPAIN	17p13.2	Amplification	FU97
SYN	22q12.3	Deletions	SCH
LRP1B	2q21.2	Focal deletions	SCH
DGKG	3q27.3	Deletions	SCH
PDE4D	5q12	Focal deletions	SCH
CCDC91	12p11.22	Deletions	MKN74
COL24A1	1q22.3	Amplification	MKN28
PAX5	9p13	Amplification	MKN28

MYO18A	17q11.2	Amplification	MKN1
STK38L	12p11.23	Focal Amplification	MKN1
SORT1	1p21.3-p13.1	Focal Amplification	MKN45
CADPS2	7q31.3	Focal Amplification	MKN45
MAP2K4	17p11.2	Deletions	MKN45
KIF18A	11p14.1	Focal Amplification	MKN45
TBC1D22A	22q13.3	Deletions	MKN45
PEPD	19q12-q13.2	Focal Amplification	TMK1
PTEN	10q23.3	Focal deletions	TMK1
LSM14A	19q13.11	Focal Amplification	TMK1
ATF7IP2	16p13.13	Amplification	YCC7
DNM2	19p13.2	Amplification	YCC7
WDR40A	9p13.3	Amplification	YCC7
ZCCHC7	9p13.2	Amplification	YCC7
MN1	22q12.1	Amplification	YCC7
FLJ33814	22q12.1	Amplification	YCC7
EMID1	22q12.2	Amplification	YCC7
NF2	22q12.2	Amplification	YCC7
RBM9	22q13.1	Amplification	YCC7
CPE	4q32.3	Deletions	SNU1
LRBA	4q31.3	Deletions	SNU5
ATE1	10q26.13	Focal Amplification	SNU16
CLCA4	1p31-p22	Deletions	YCC2
MEP1A	6p12-p11	Focal Amplification	YCC2

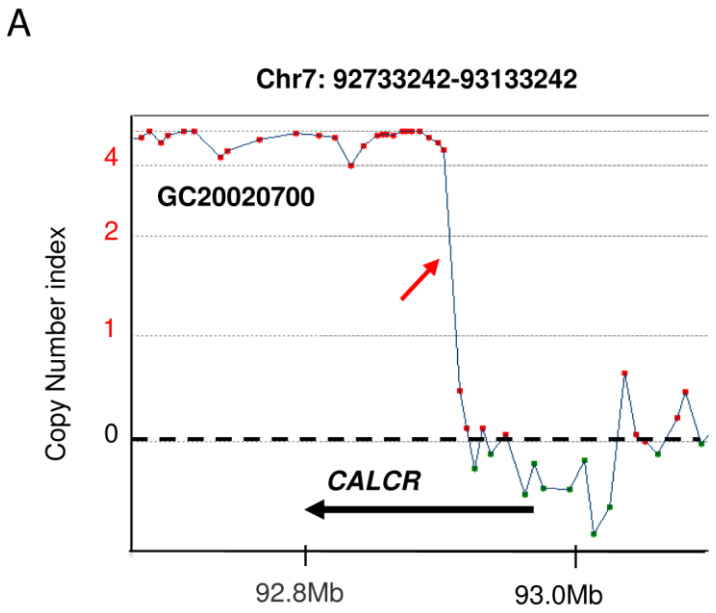


Figure 3.2. Genomic breakpoint analysis in GC

A. Representative example of a genomic breakpoint. aCGH profile of GC tumor GC20020700 exhibited a genomic breakpoint in the *CALCR* gene on Chr 7q12. *X-axis* - physical chromosomal coordinates. *Y-axis* - log2-transformed smoothened values (i.e. 0 indicates copy number equals to 2). The red arrow indicates the breakpoint of interest. The black arrow below represents the *CALCR* gene pointing in a 5'-3' direction. Each dot represents a microarray probe.

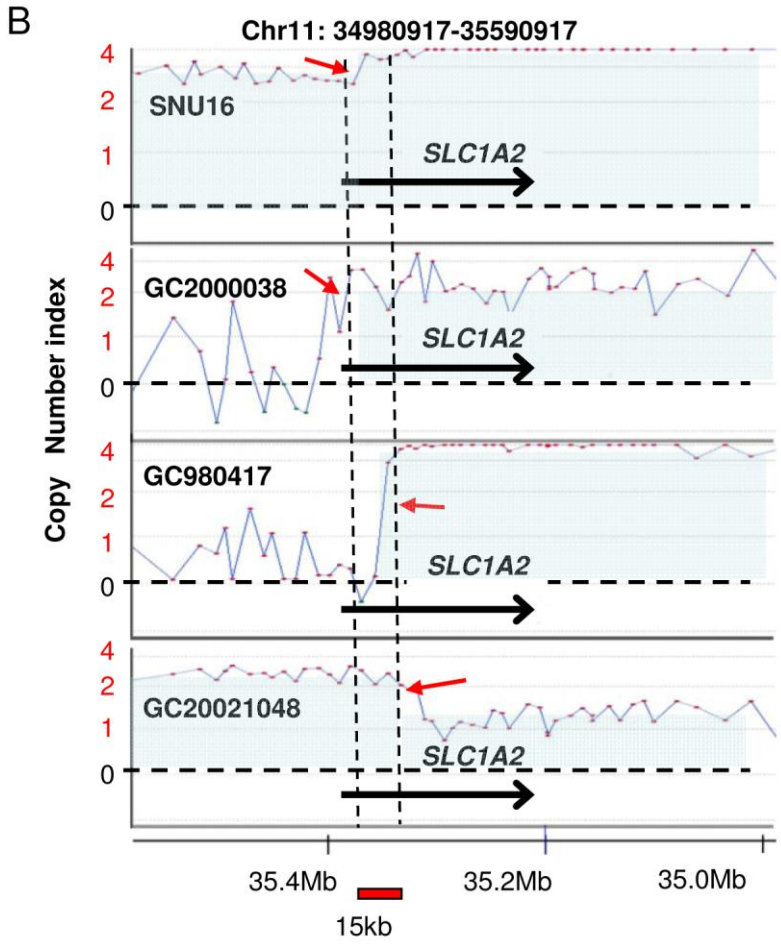


Figure 3.2. Genomic breakpoint analysis in GC

B. Genomic breakpoint analysis in GC identified recurrent breakpoints in the 5' region of *SLC1A2* in four GCs (3 primary tumors and 1 cell line - GC2000038, GC980417, GC20021048, SNU16). Red arrows indicate the breakpoints of interest. *X-axis* - physical chromosomal coordinates. *Y-axis* - log2-transformed smoothed values (i.e. 0 indicates copy number equals to 2). The black arrow below represents the *SLC1A2* gene pointing in a 5'-3' direction. Each dot represents a microarray probe. The red bar below indicates the minimal common recurrent breakpoint region of 15kb.

3.1.3 Validation of *SLC1A2* genomic breakpoints

Notably, 11p13-15 where *SLC1A2* are located has been described as a frequent site of genome rearrangement in gastric and esophageal cancers (Rodriguez et al., 1990). Therefore, we performed Spectral Karyotyping (SKY) to identify possible genomic rearrangements in SNU16. SKY analysis confirmed the presence of at least two 11p13-11p14 genome rearrangements in SNU16 cells, one involving fusion of Chromosome 1 with Chromosome 11 at band 11p13-14, and the second involving a complex chromosomal scenario with rearrangements joining chromosomes 5, 10 and 11 (Fig. 3.3A). These results confirmed previous findings that 11p13-15 is a frequent site for genomic aberrations.

To validate the *SLC1A2* breakpoint region, we performed Fluorescence *in situ* Hybridization (FISH) analysis using fosmid probes mapping upstream or downstream to the putative breakpoint (WI2-67O19 and WI2-1928P9). Supporting the aCGH data, the WI2-67O19 upstream probe (35384118-35427600) covering the first exon of *SLC1A2* showed 4 signals in SNU16 nuclei (Fig. 3.3C, left), confirming previous studies which reported SNU16 is a naturally tetraploid cell line (Park et al., 1990). In contrast, the downstream WI2-1928P9 probe (35323126-35359663) located at *SLC1A2* intron 1 showed multiple hybridization signals (>50 copies) indicating a high level amplification event (Fig. 3.3C, right).

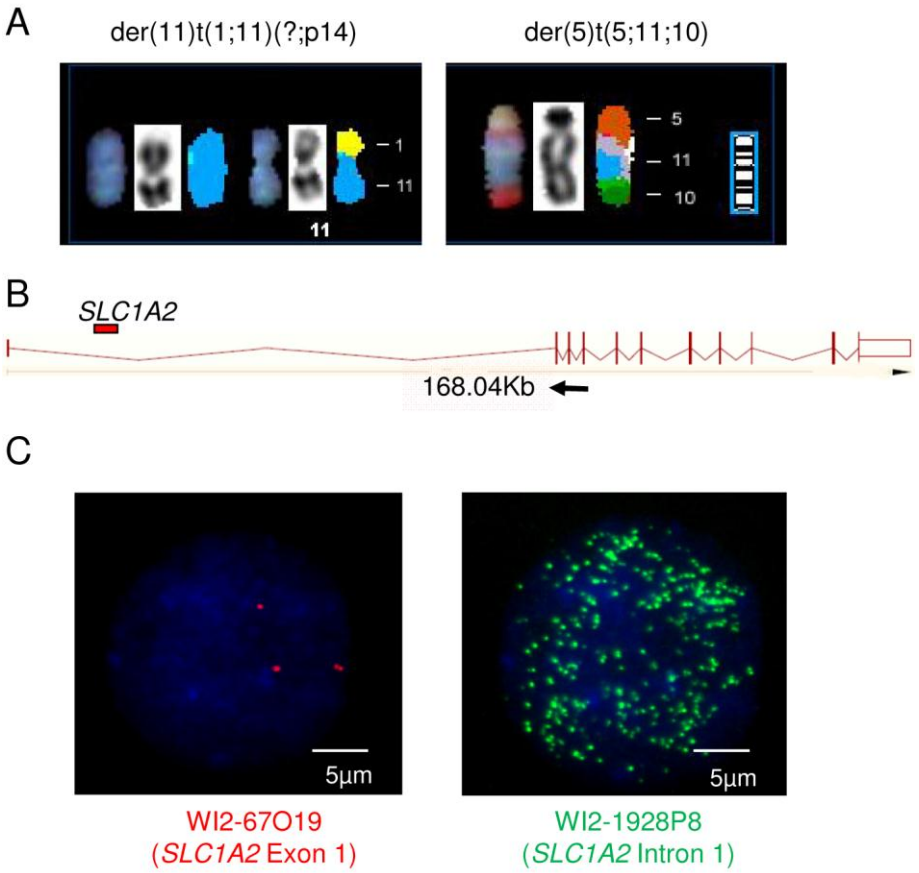


Figure 3.3. Validation of unbalanced genomic break in 11p13 *SLC1A2* region in SNU16 cells

A. SKY analysis of SNU16 cells revealed an unbalanced chromosomal rearrangement affecting 11p13-11p14. *Left*, SKY showed chromosome aberration t(1;11)(?;p14), indicating a rearrangement of chromosome 1 with chromosome 11 at bands 11p14. *Right*, SKY identified the complex chromosomal aberration t(5;11;10) with rearrangements joining material from three different chromosome 5, 11, 10.

B. Genomic organization of the *SLC1A2* gene. Vertical bars represent *SLC1A2* exons, connected by intervening introns. Total length of the *SLC1A2* gene is 168 Kb. Red bar: minimal common recurrent breakpoint region in *SLC1A2* intron 1 (15-24 kb).

C. FISH Validation of *SLC1A2* breakpoints. Probes WI2-67O19 (red) covered *SLC1A2* exon 1; probes WI2-1928P9 (green) covered *SLC1A*2 intron 1 region.

3.2 *SLC1A2* breakpoint characterization reveals a *CD44-SLC1A2* gene fusion

3.2.1 Identify 5' fusion partners to *SLC1A2*

Integrating the *SLC1A2* breakpoint regions from the aCGH and the FISH data, we defined a 15-24 kb minimal common breakpoint window in the *SLC1A2* first intron (Figure 3.2B and 3.3B, red bar). We hypothesized that chromosomal aberrations affecting this region may disrupt the *SLC1A2* gene resulting in potential fusion partners. To test this possibility, we performed 5' RNA Ligase-Mediated Rapid Amplification of cDNA Ends (RLM-RACE) to characterize *SLC1A2* transcript sequences upstream to *SLC1A2* exon 2. A 250 bp 5' RACE product was identified in breakpoint positive SNU16 cells, but not in other GC cell lines without *SLC1A2* breakpoints (AGS, YCC1, YCC9 and N87) (Fig. 3.4A).

Sequencing of the amplified SNU16 RACE product revealed a *CD44-SLC1A2* fusion transcript, formed by the juxtaposition of *CD44* exon 1 to *SLC1A2* exon 2 (Fig. 3.4B). To validate the 5' RACE results, we designed combination sets of PCR primers targeting *CD44* exon 1 (forward primer) and *SLC1A2* exons 3, 4, 5, 6 (reverse primers) to directly detect the fusion by reverse-transcription PCR (RT-PCR). *CD44-SLC1A2* transcripts were detected in SNU16 cells, but not in other GCCLs nor in commercially available normal gastric tissue (NG) (Fig. 3.4C-D). These results demonstrate the existence of a novel *CD44-SLC1A2* fusion transcript in SNU16 cells.

We further confirmed the expression of a complete ~1.6 kb *CD44-SLC1A2* transcript in SNU16 cells using RT-PCR primers targeting *CD44* exon 1 and *SLC1A2* exon 11 (the last *SLC1A2* exon) (Fig. 3.4E), with two potential translation start sites (Fig. 3.4B, underlined) .

All of the SLC family gene fusions identified to date involve the *SLC* gene at the 5' end of the fusion, such as *SLC45A3-BRAF* in prostate cancer and *SLC34A2-ROS1* in non-small-cell lung cancer. So in complement to the 5' RACE analysis, we performed 3' RACE from exon 1 of *SLC1A2* to determine if there is a productive 3' partner for *SLC1A2*. However, 3' RACE analysis in SNU16 cells studying transcripts downstream of *SLC1A2* exon 1 did not identify any additional fusion partners besides wild-type *SLC1A2* transcripts (Fig. 3.4F). Thus, at this point, we have no evidence suggesting that the *SLC1A2* 5' end is involved in another productive fusion in SNU16.

CD44 and *SLC1A2* lie adjacent to each other on chromosome 11p13, being separated by only ~19kb (Fig. 3.4G). The two genes are transcribed towards each other and lie on opposite strands, indicating that they possess distinct promoters. Because chimeric transcripts caused by read-through transcription typically involve adjacent genes on the same strand (Akiva et al., 2006; Wang et al., 2007), it is unlikely that the *CD44-SLC1A2* fusion transcripts are created by a read-through event. We thus hypothesized that the *CD44-SLC1A2* gene fusion might have been caused by a paracentric chromosomal inversion.

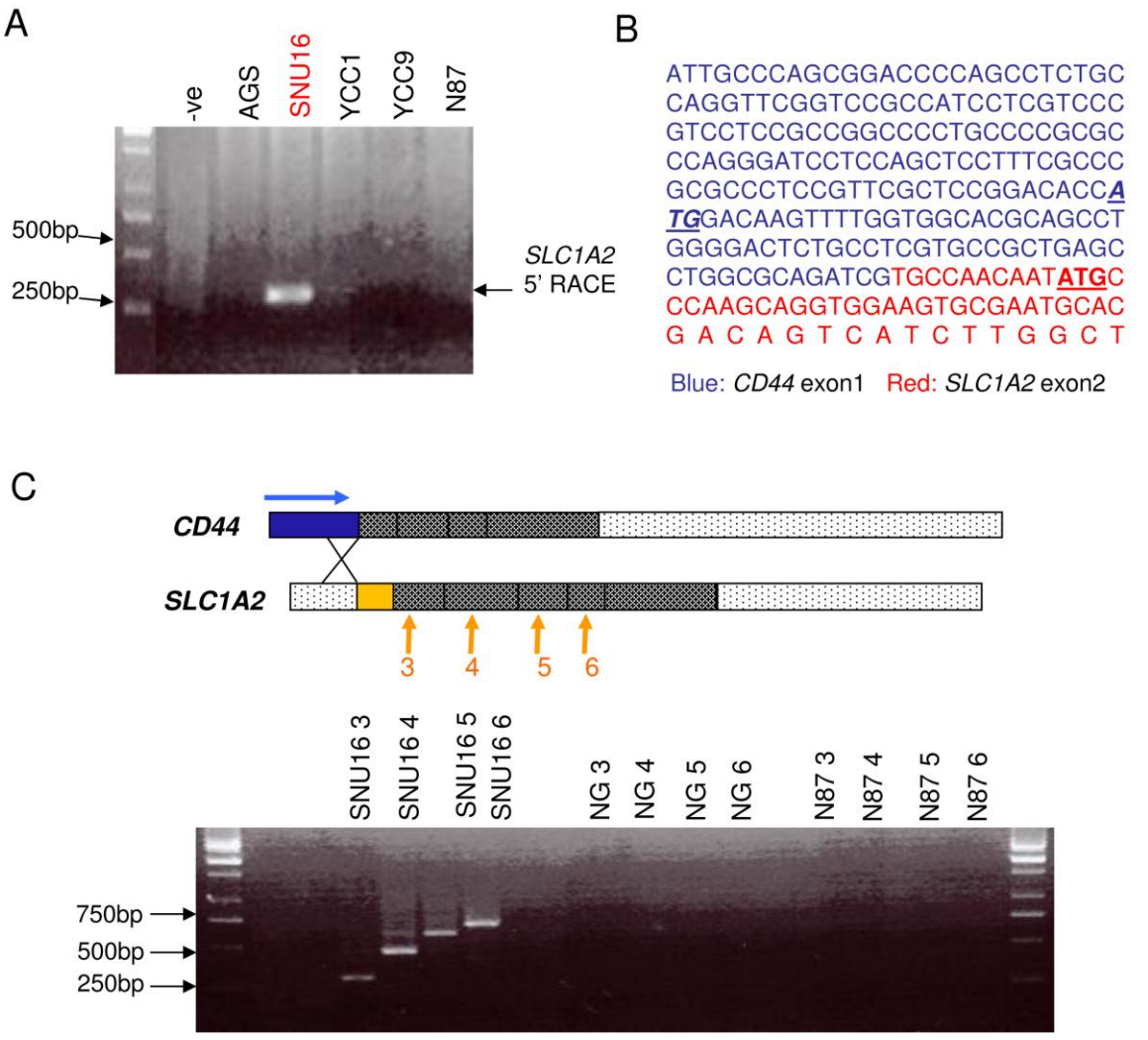


Figure 3.4. Detection of fusion *CD44-SLC1A2* in SNU16 cell line

A. 5' *SLC1A2* RLM-RACE in GC cell lines (AGS, SNU16, YCC1, YCC9, N87). PCR mix without sample input was used as a negative control (-ve).

B. Sequencing of the 5' RACE gene product identified a fusion of *CD44* exon 1 to *SLC1A2* exon 2. Blue colored sequence indicates the *CD44* exon 1. Red sequence shows partial exon 2 of *SLC1A2*. ATG sites are underlined.

C. RT-PCR confirmed the exon joining between *CD44* and *SLC1A2* in SNU16 cells. 4 RT-PCR reactions were conducted using a forward primer in exon 1 of *CD44* indicated by the blue horizontal arrow, and 4 different reverse primers in exons 3, 4, 5, 6 of *SLC1A2* respectively (orange arrows). NG – normal stomach, N87 – fusion negative line.

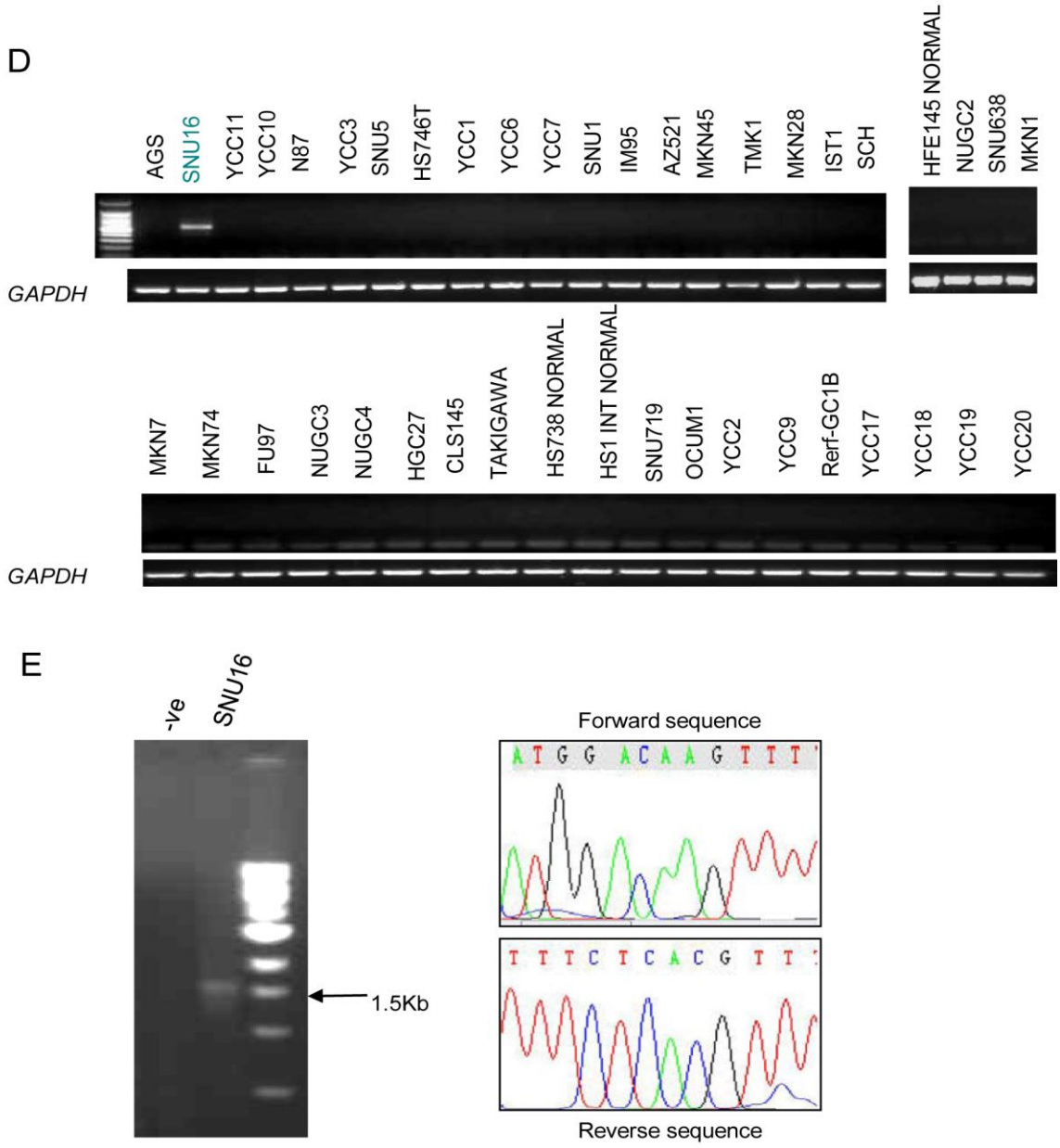


Figure 3.4. Detection of fusion *CD44-SLC1A2* in SNU16 cell line

D. RT-PCR screening of the *CD44-SLC1A2* fusion in 42 gastric cancer cell lines and gastric normal cell lines. SNU16 (green) is the only cell line expressing the fusion.

E. *Left*, RT-PCR analysis using a forward primer to *CD44* exon 1 and reverse primer to *SLC1A2* exon 11 (last exon) confirmed expression of a ~1.6 kb *CD44-SLC1A2* fusion transcript in SNU16 cells. *Right*, PCR product sequencing confirmed a full length fusion in SNU16.

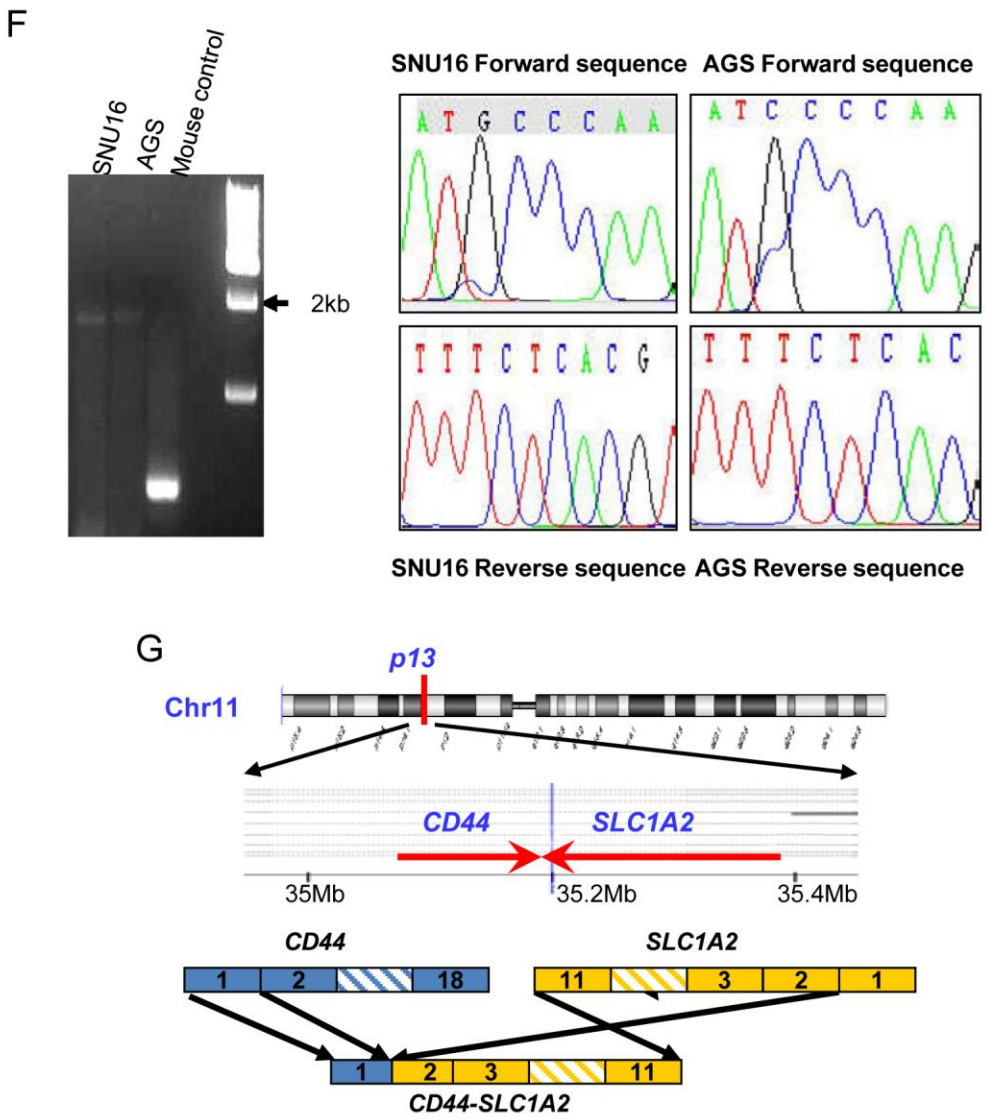


Figure 3.4. Detection of fusion *CD44-SLC1A2* in SNU16 cell line

F. *Left*, 3'RACE analysis of *SLC1A2* identified 2kb products in both AGS and SNU16. *Right*, sequencing of the 3' RACE products confirmed wild-type full length *SLC1A2* transcripts in both SNU16 and AGS cells.

G. Predicted fusion pattern between *CD44* and *SLC1A2*. *Top*, *CD44* and *SLC1A2* chromosomal organization. *Bottom*, *CD44-SLC1A2* relationship to *CD44* and *SLC1A2* parent genes. Blue boxes represent *CD44* exons; Orange boxes represent *SLC1A2* exons.

3.2.2 Confirmation of *CD44-SLC1A2* chromosomal inversion in SNU16

The close proximity of the *CD44* and *SLC1A2* genes makes it extremely challenging to validate the *CD44/SLC1A2* genomic inversion using conventional FISH techniques, which have typical resolutions of 100 kb - 1 Mb. In the current study, we used two different strategies to verify the presence of *CD44/SLC1A2* genomic inversions in this region. First, we used fiber-FISH, a high-resolution method for genomic DNA mapping (Florijn et al., 1995). Fosmid probes Rp1-68d18 (35146316-35329998, covering the *CD44* gene and a portion of the *SLC1A2* gene, green signal) and Rp11-1148l23 (35294107-35461767, covering the *SLC1A2* gene only, red signal) were hybridized to SNU16 cells or normal lymphoblastoid CCL159 cells. In control CCL159 cells, a normal chromosome was observed and indicated by two distinct red and green probe signals lying adjacent to one another. In contrast, a "split-apart" red-green-red signal was detected in SNU16 cells, consistent with an inversion event occurring between these probes (Fig. 3.5A).

Second, we directly confirmed a *CD44/SLC1A2* genomic inversion in SNU16 cells using long-range genomic PCR, followed by end-sequencing of the PCR products. Using primers located to *CD44* exon 1 and the *SLC1A2* first intron (black arrows in Fig. 3.5A), we successfully PCR amplified and sequence validated a *CD44/SLC1A2* inversion product in SNU16 fusion-positive cells but not in AGS cells (Fig. 3.5B). Taken collectively, these two alternative methods confirm the presence of a chromosomal inversion event in SNU16 occurring between *CD44* and *SLC1A2*.

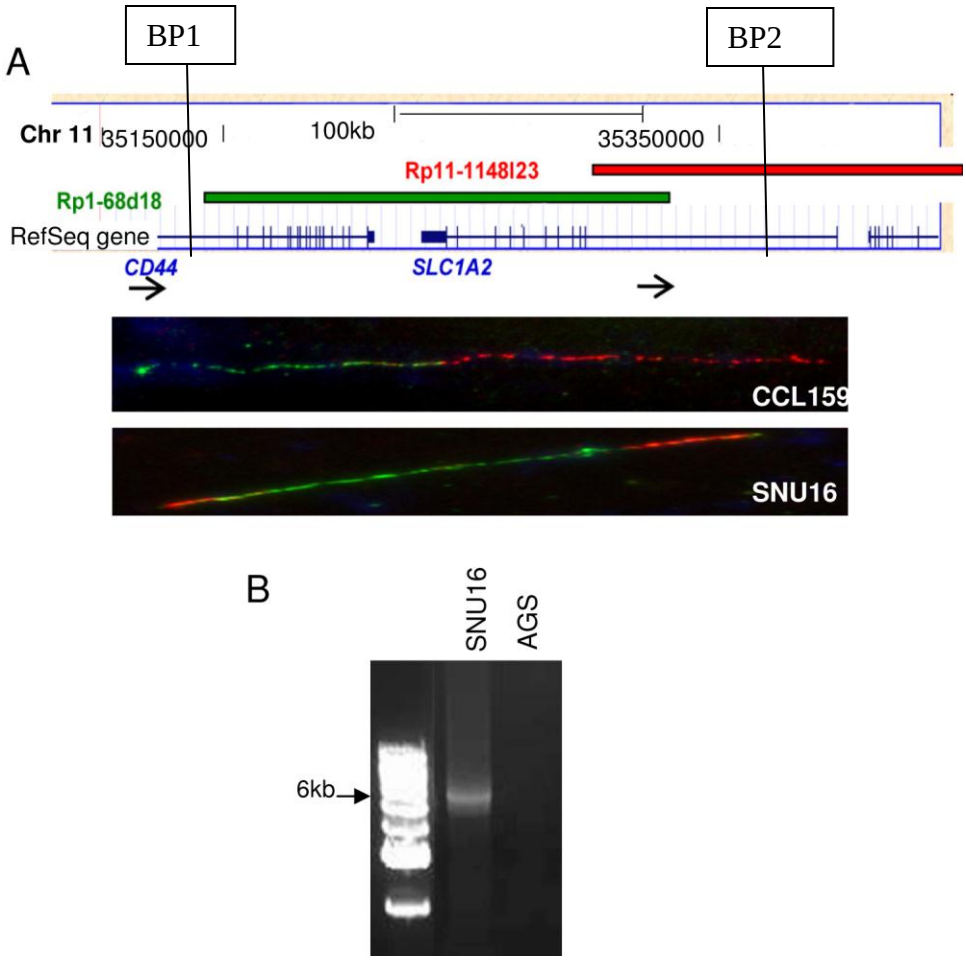


Figure 3.5. Confirmation of chromosomal inversion model of *CD44-SLC1A2* gene fusion in SNU16

A. Fiber-FISH analysis revealed an inversion of *SLC1A2* to *CD44* in SNU16 cells. *Top*, Rp1-68d18 biot covers *CD44* (3' of intron 1) and *SLC1A2* (3' of intron 1). Rp11-1148l23 dig covers the 5' region of *SLC1A2* intron 1 and upstream sequence. *Bottom*, Fiber-FISH images of control lymphoblast cell line CCL159 and fusion-positive SNU16 cells.

B. Long-range PCR confirmed a chromosomal inversion of *CD44-SLC1A2* in SNU16. Primers were targeted to *CD44* exon 1 and the *SLC1A2* first intron (black arrows in (A)). SNU16 – fusion positive GC cell line. AGS – fusion negative GC cell line.

3.2.3 *CD44-SLC1A2* protein expression

Sequence analysis of the *CD44-SLC1A2* fusion revealed two distinct protein translation patterns (Fig. 3.6A). First, translation initiating from an ATG site in *CD44* exon 1 could produce a 65 aa protein, comprising 22 aa of *CD44* and 43 aa of novel sequence. Second, protein translation might also initiate from an alternative ATG site in *SLC1A2* exon 2, downstream of the fusion site. Notably, alternative *SLC1A2* splice forms initiating from *SLC1A2* exon 2 have been reported (see Ensembl EST database ENST00000395753/ENSP00000379102), consistent with protein translation initiating from the exon 2 ATG. Translation from this alternative ATG would produce a 565 amino acid truncated *SLC1A2* protein that is 9 amino acids shorter than the full length form, but retaining all functionally relevant domains including transmembrane helices and symporter domains.

We first conducted immunofluorescence assays on fusion positive SNU16 cells and fusion negative AGS cells using SLC1A2 antibodies. SNU16 cells expressed strong immunoreactivity to *SLC1A2* antibodies in a strong membranous pattern. However, fusion negative AGS cells expressed cytoplasmic SLC1A2 (Fig. 3.6B). These results suggested that the *CD44-SLC1A2* fusion might produce a SLC1A2 protein that is mostly localized in the membrane.

To test if *CD44-SLC1A2* produces a truncated SLC1A2 protein, we performed western blotting using anti-SLC1A2 antibodies on fusion positive and fusion negative GC cells. We specifically isolated proteins from the cell membrane. In SNU16 fusion-positive cells, we detected a smaller-sized SLC1A2 protein compared to fusion negative AGS and

SNU5 cells (Fig. 3.6C), consistent with translation initiating from *SLC1A2* exon 2 in SNU16 cells. To further demonstrate that the alternative ATG in *SLC1A2* exon 2 is capable of initiating protein translation, we cloned and expressed the full-length *CD44-SLC1A2* fusion gene in HFE145 gastric normal epithelial cells (Akhtar et al., 2001). Western blotting analysis confirmed expression of an immunoreactive SLC1A2 product in *CD44-SLC1A2* transfected HFE145 cells of the expected size (Fig. 3.6D). This demonstrates that the alternative ATG in *SLC1A2* exon 2 is sufficient to initiate translation. These findings suggest that the biological function of *CD44-SLC1A2* is likely to be complex and its specific role in cancer development remains to be determined.

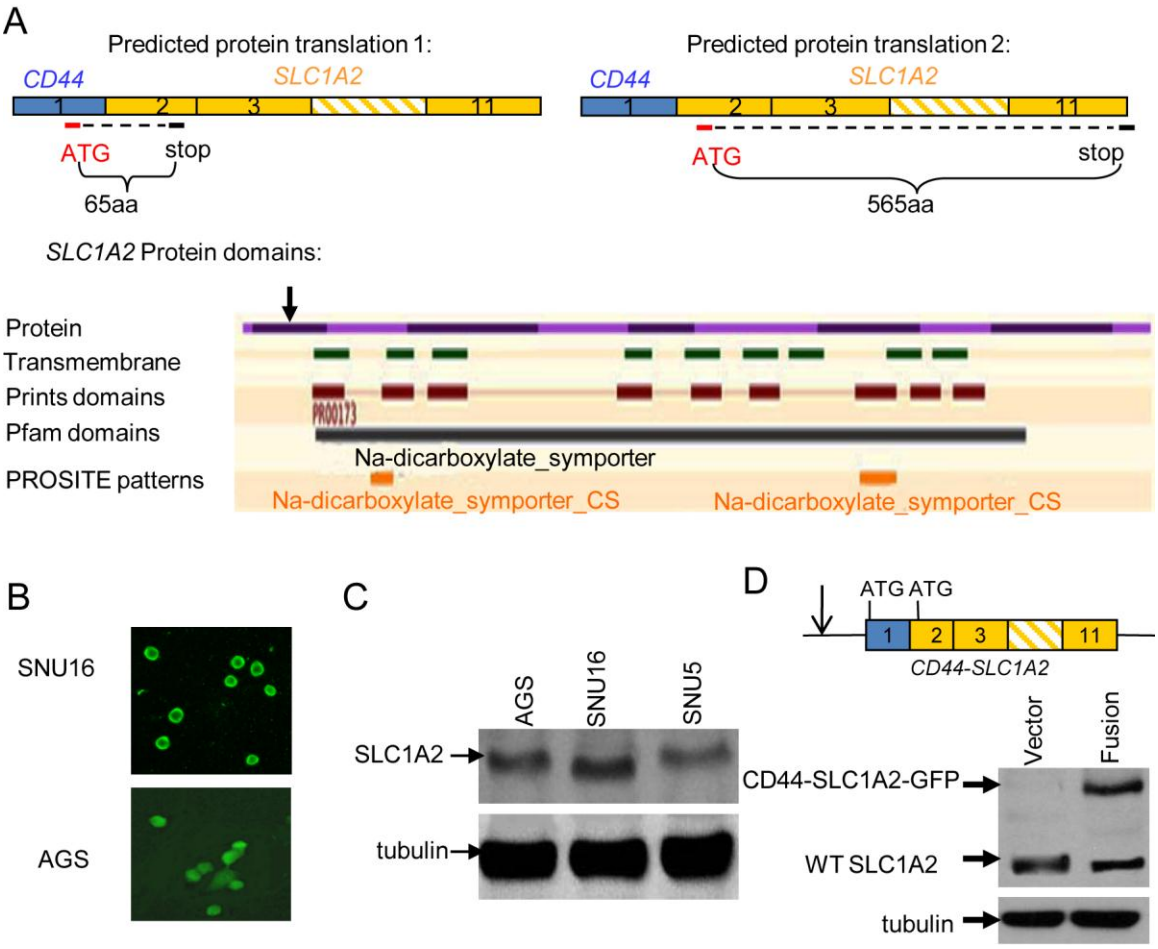


Figure 3.6. Protein expression of *CD44-SLC1A2*

A. Predicted *CD44-SLC1A2* protein translation patterns. *Top left*, translation initiates from an ATG site in *CD44* exon 1, producing a 65aa protein. *Top right*, translation initiates from an alternative ATG site in *SLC1A2* exon 2, producing a truncated *SLC1A2* protein of 565aa. *Bottom*, diagram illustrates the known protein domains of full length *SLC1A2*. The black arrow indicates the position of second protein translation initiation site. A truncated SLC1A2 protein beginning from exon 2 is predicted to encode all known functional domains.

B. Immunofluorescence assays on fusion positive SNU16 cells and fusion negative AGS cells using SLC1A2 antibodies. SNU16 cells expressed strong membranous SLC1A2 expression. AGS cells expressed endogeneous SLC1A2.

C. Western blot of fusion-positive SNU16 and fusion-negative AGS and SNU5 cells (membrane fractions). *Top,* anti-SLC1A2 antibodies. *Bottom,* α-tubulin control.

D. *CD44-SLC1A2* ectopic expression in HFE145 normal gastric epithelial cell line. *Top, CD44-SLC1A2* expression construct carrying a GFP tag. Arrow – promoter. ATG sites in *CD44* exon 1 and *SLC1A2* exon 2 are shown. *Bottom,* Immunoblotting with anti-SLC1A2 antibodies. Vector-HFE145 cells transfected with vector control; Fusion- HFE145 cells transfected with *CD44-SLC1A2* expression construct.

3.3 Functional analysis of *CD44-SLC1A2* fusion in GC cells

3.3.1 Efficacy of the fusion-specific siRNA1

To investigate the functional consequences of inhibiting *CD44-SLC1A2* expression, we designed two customized siRNAs targeting the *CD44-SLC1A2* fusion site to avoid confounding influences from wild-type *CD44* and *SLC1A2* transcripts. Treatment of SNU16 cells with 100 nM fusion specific siRNA1 successfully silenced *CD44-SLC1A2* expression after 48 hours, but did not discernibly alter the independent expression of *CD44* or *SLC1A2* (Fig. 3.7A). To confirm silencing at the protein level, we conducted immunofluorescence assays on both control and silenced SNU16 cells using SLC1A2 antibodies. Untreated SNU16 cells expressed strong immunoreactivity to SLC1A2 antibodies in a strong membranous pattern, which was significantly reduced by siRNA treatment (Fig. 3.7B). We further confirmed silencing at the protein level by western blotting (Fig. 3.7C). These results confirm the efficacy of the fusion-specific siRNA1.

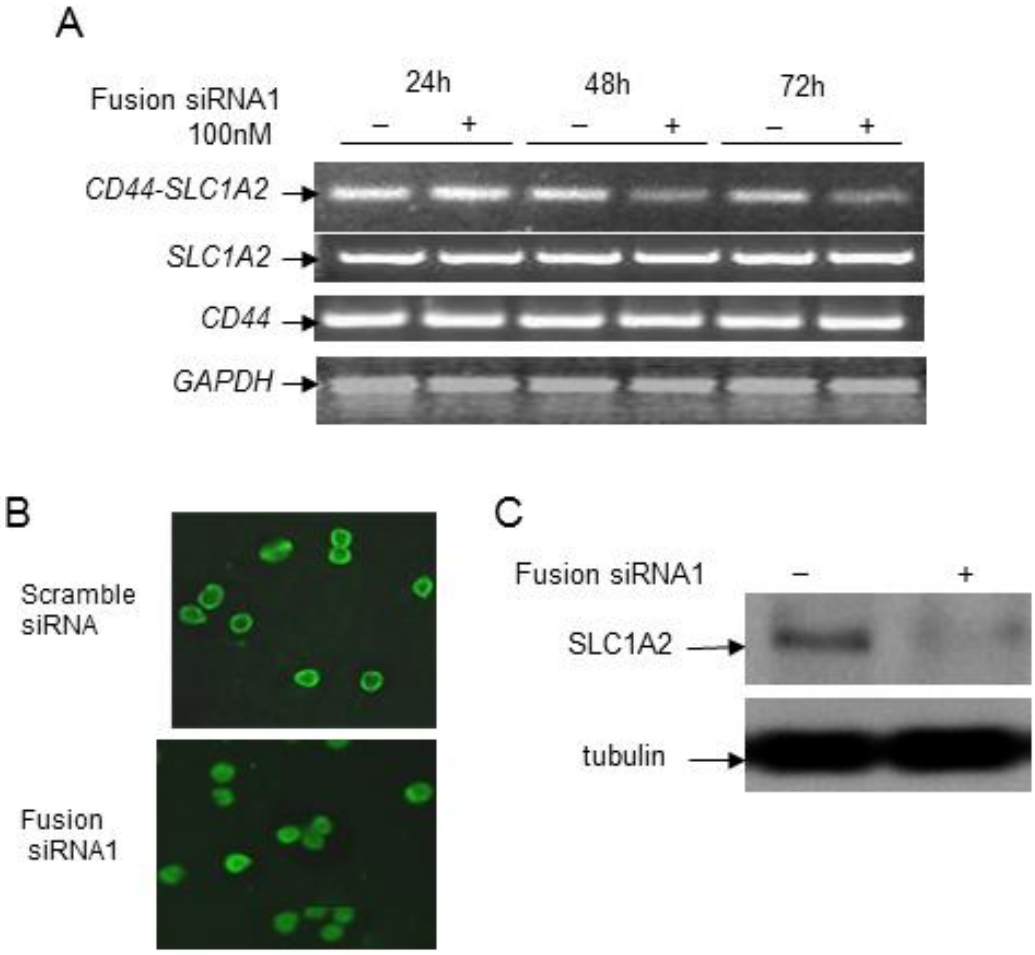


Figure 3.7. *CD44-SLC1A2* silencing by fusion-specific siRNA1 inhibits cellular proliferation, colony formation and invasion in SNU16

A. *CD44-SLC1A2* silencing by fusion-specific siRNA1 (CGCAGAUCGU GCCAACAAUUU) in SNU16. *CD44-SLC1A2* expression was measured after 24h, 48h and 72h post siRNA treatment. *CD44*: wild-type *CD44* expression. *CD44* primers were designed to target exons 3-5. *SLC1A2*: wild-type *SLC1A2* expression. *SLC1A2* primers were designed to target exon 1. *GADPH* was used as a loading control.

B. Immunofluorescence assays on both control and silenced SNU16 cells using SLC1A2 antibodies. Untreated SNU16 cells expressed strong membranous expression of SLC1A2, which was significantly reduced by siRNA1 treatment.

C. Western blotting in SNU16 with *CD44-SLC1A2* silencing by *CD44-SLC1A2* siRNA1. SLC1A2 protein levels were monitored using anti-SLC1A2 antibodies. Tubulin was used as a loading control. – and +: SNU16 cells transfected with scramble siRNA (-) and fusion-specific siRNA1 (+).

3.3.2 *CD44-SLC1A2* silencing using siRNA1 reduces cancer cell proliferation, invasion, and colony formation

We first silenced *CD44-SLC1A2* in SNU16 and monitored proliferation changes. We observed a significant reduction in cell proliferation capacity upon siRNA1 treatment in SNU16 cells (Figure 3.7D, p=0.002). This result suggests that *CD44-SLC1A2* may be important for cancer cell proliferation in GC. We also performed colony formation assay to assess the tumorigenicity of SNU16 upon *CD44-SLC1A2* knockdown. We observed a potent inhibition of colony formation in fusion-silenced cells compared to controls (p=0.01, Figure 3.7E). We further conducted matrigel assays to investigate effects of *CD44-SLC1A2* on cancer cell invasion. *CD44-SLC1A2* -silenced SNU16 cells exhibited a decreased level of cell invasion compared with control cells (Figure 3.7F, p=0.0013), suggesting a potential role for *CD44-SLC1A2* in cell motility and invasion.

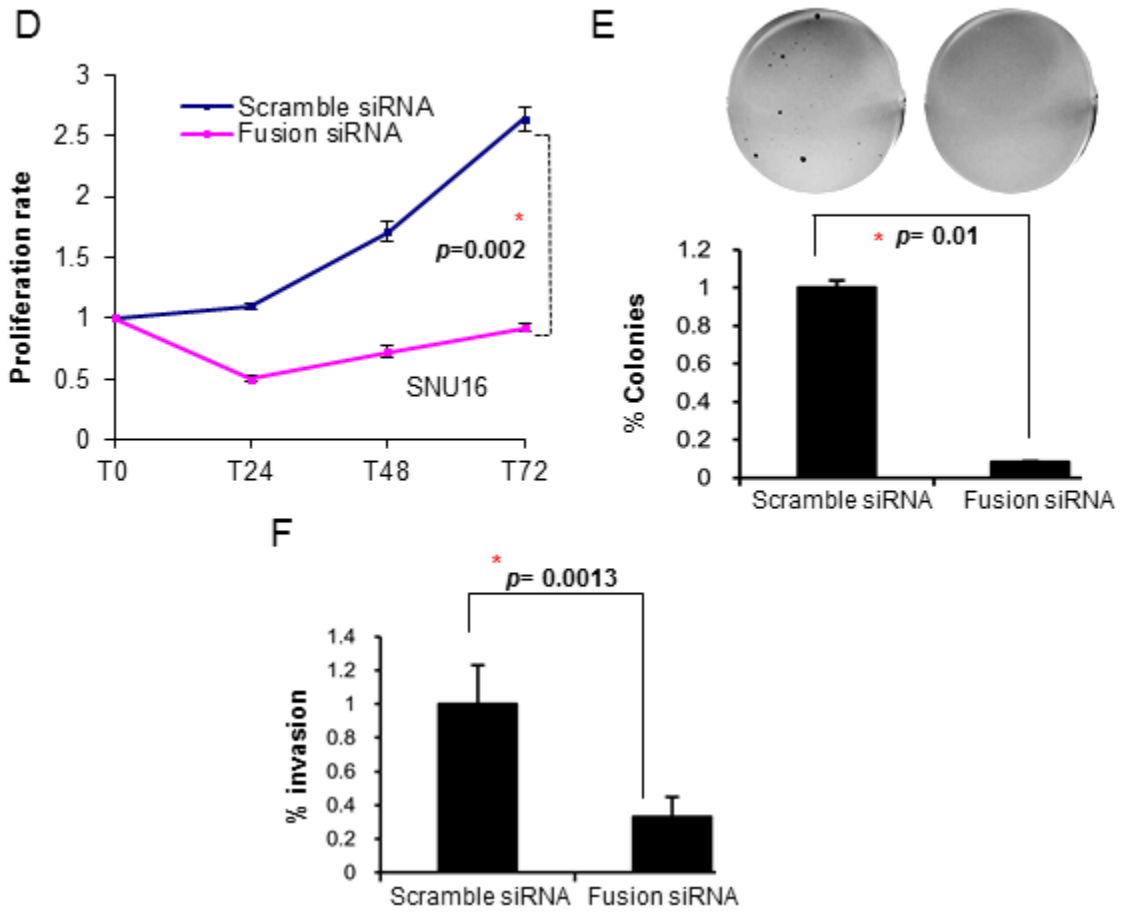


Figure 3.7. *CD44-SLC1A2* silencing by fusion-specific siRNA1 inhibits cellular proliferation, colony formation and invasion in SNU16

D-F. Functional effects of *CD44-SLC1A2* silencing by fusion-specific siRNA1 in SNU16.

D. SNU16 cells treated with *CD44-SLC1A2* siRNA1 exhibited a significant decrease in proliferation rate (p=0.002) compared to scramble siRNA treated cells. Proliferation rates were measured at 24h, 48h and 72h post siRNA treatment. P value was calculated at T72 using student's unpaired t-test.

E. *CD44-SLC1A2* silencing by fusion-specific siRNA1 significantly reduced tumorogenicity compared to scramble siRNA treated cells (p=0.01). P value was calculated using student's unpaired t-test.

F. Cell invasion assays using SNU16 cells treated with scramble siRNA and *CD44-SLC1A2* siRNA1. Knockdown of *CD44-SLC1A2* resulted in a significant reduction of invasion rate (p=0.0013) compared to scramble siRNA treated cells. P value was calculated using student's unpaired t-test.

3.3.3 Efficacy of the fusion-specific siRNA2

Due to the need for targeting a specific fusion site sequence, it was not possible to design a completely non-overlapping siRNA to confirm these results. However, when we used a second *CD44-SLC1A2* targeting siRNA containing overlapping but distinct sequence, similar results were obtained as compared to fusion specific siRNA1. Treatment of SNU16 cells with 100 nM fusion specific siRNA2 successfully silenced *CD44-SLC1A2* expression after 72 hours, but did not alter the independent expression of *CD44* or *SLC1A2* (Fig. 3.8A). We also conducted immunofluorescence assays on both control and silenced SNU16 cells. Similar to fusion specific siRNA1 silencing, SNU16 cells transfected with scramble siRNA expressed strong immunoreactivity to SLC1A2 antibodies in membrane, which was significantly reduced by siRNA treatment (Fig. 3.8B). Western blotting also confirmed silencing at the protein level by siRNA2 (Fig. 3.8C), suggesting the efficacy of the fusion-specific siRNA2.

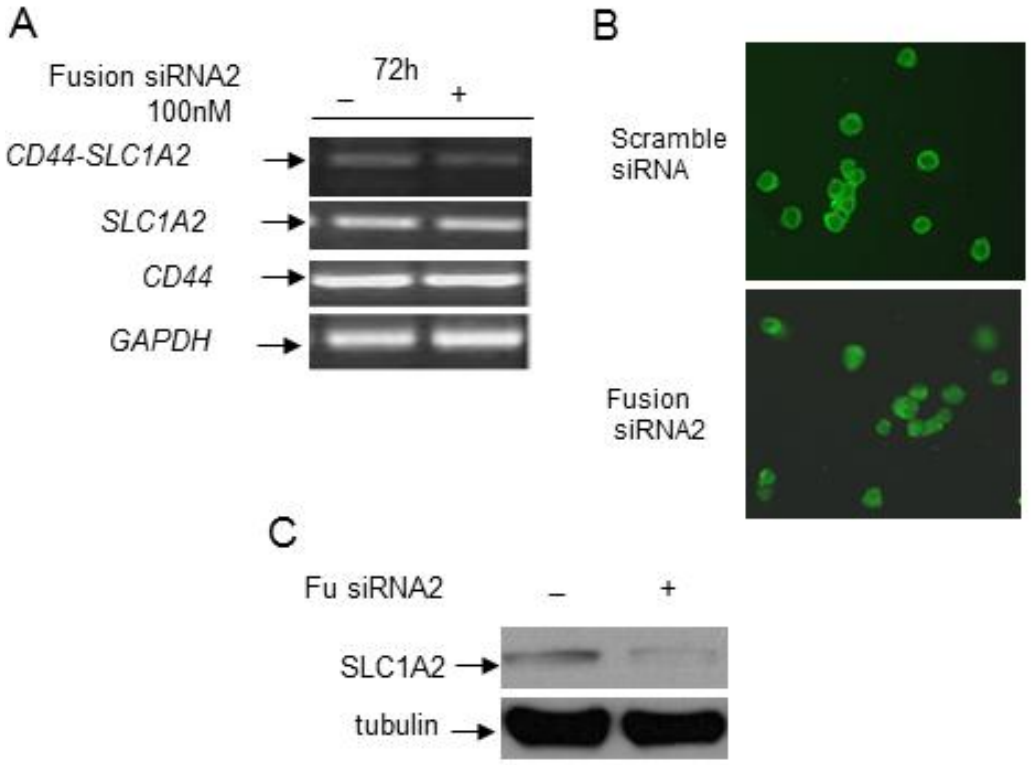


Figure 3.8. Silencing *CD44-SLC1A2* with a second fusion specific siRNA inhibits cellular proliferation, invasion, and colony formation in SNU16

A. Silencing of *CD44-SLC1A2* by fusion-specific siRNA2 (GCACAUCGUGCC AACAAUAUU). Reduction of *CD44-SLC1A2* expression was observed after siRNA2 treatment, without impairing regular *CD44* and *SLC1A2* expression. *CD44*: wild-type *CD44* expression. *CD44* primers were designed to target exons 3-5. *SLC1A2*: wild-type *SLC1A2* expression. *SLC1A2* primers were designed to target exon 1. *GADPH* was used as a loading control.

B. Immunofluorescence assays on both control and silenced SNU16 cells using SLC1A2 antibodies. Untreated SNU16 cells expressed strong membranous expression of SLC1A2, which was significantly reduced by siRNA2 treatment.

C. Western blotting analysis confirms knockdown of SLC1A2 protein expression in SNU16 (membrane fraction). SLC1A2 protein levels were monitored using anti-SLC1A2 antibodies. Tubulin control confirmed equal amount of protein loading in both lanes. – and +: SNU16 cells transfected with scramble siRNA (-) and fusion-specific siRNA2 (+).

3.3.4 *CD44-SLC1A2* silencing using siRNA2 reduces cancer cell proliferation, invasion, and colony formation

Similar to siRNA1 silencing, we observed a significant reduction in cell proliferation capacity upon siRNA2 treatment in SNU16 cells (Figure 3.8D, $p<0.001$), further suggesting that *CD44-SLC1A2* may be important for cancer cell proliferation in GC. In addition, we observed a potent inhibition of colony formation in fusion-silenced cells compared to controls ($p<0.001$, Figure 3.8E). We also conducted matrigel assays to investigate effects of *CD44-SLC1A2* on cancer cell invasion. *CD44-SLC1A2* silenced SNU16 cells exhibited a decreased level of cell invasion compared with control cells (Figure 3.8F, $p=0.04$), suggesting a potential role for *CD44-SLC1A2* in cell motility and invasion. Taken collectively, these results support a pro-oncogenic role for *CD44-SLC1A2* in cancer cell proliferation, invasion and colony formation.

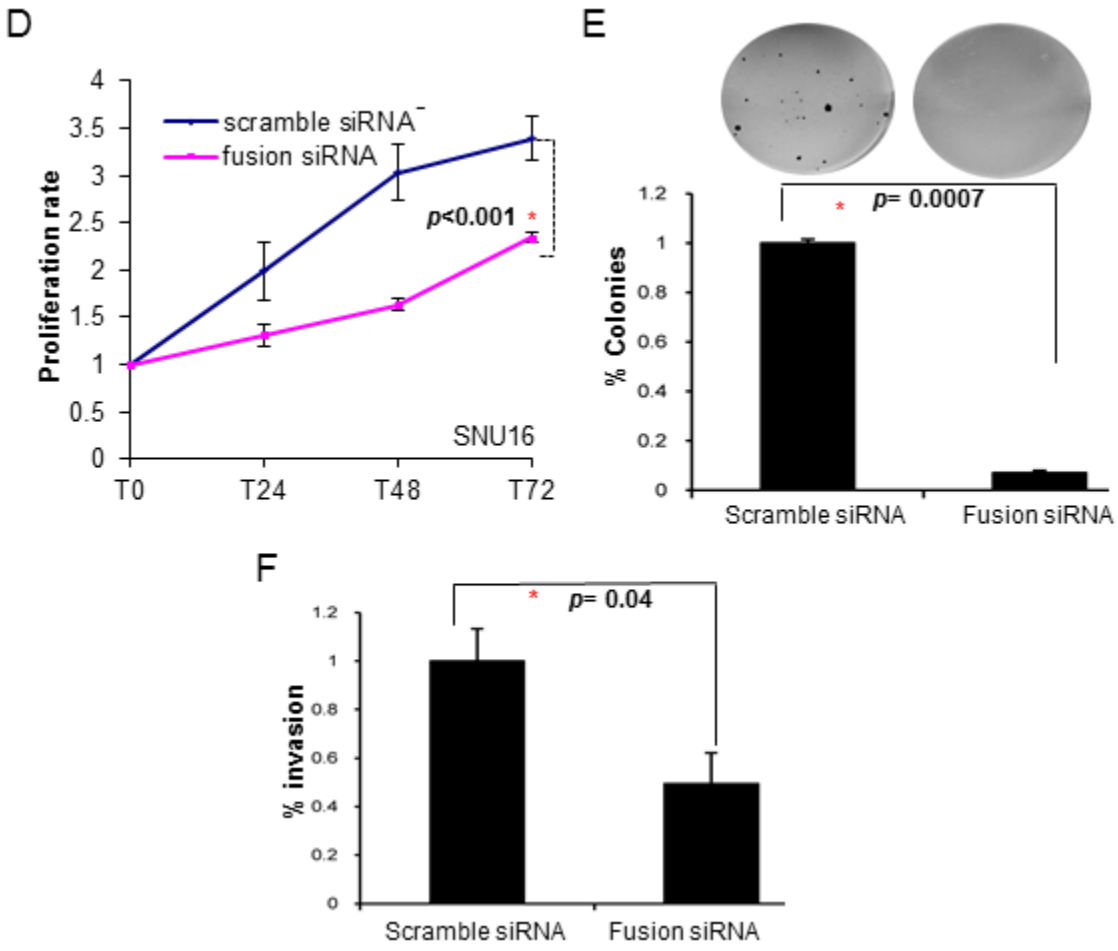


Figure 3.8. Silencing *CD44-SLC1A2* with a second fusion specific siRNA inhibits cellular proliferation, invasion, and colony formation in SNU16

D-F. Functional effects of *CD44-SLC1A2* silencing by fusion-specific siRNA2 in SNU16.

D. SNU16 cells treated with *CD44-SLC1A2* siRNA2 exhibit a significant decrease in proliferation rate ($p<0.001$) compared to scramble siRNA treated cells. Proliferation rates were measured at 24h, 48h and 72h post siRNA treatment. P value was calculated at T72 using student's unpaired t-test.

E. Colony formation assays revealed a reduction in numbers of colonies formed by fusion specific siRNA2 treatment compared to scrambled siRNA treatment (p=0.0007). P value was calculated using student's unpaired t-test.

F. Knockdown of *CD44-SLC1A2* using fusion siRNA2 also resulted in a significant reduction of invasion rate ($p=0.04$). P value was calculated using student's unpaired t-test.

3.3.5 Fusion specific siRNAs knockdown in fusion negative AGS cells

To exclude the possibility of off target effects of fusion specific siRNAs, we transfected fusion negative AGS cells with the two customized fusion siRNAs. As expected, we observed no change in the expression of the wild type *CD44* or *SLC1A2* (Fig. 3.9A-B, left). In addition, no significant change in proliferation rate was observed after fusion specific siRNAs knockdown in AGS cells (Fig. 3.9A-B, right).

3.3.6 Wild type *SLC1A2* siRNAs knockdown in fusion negative AGS cells

The observation that *CD44-SLC1A2* produces an almost full-length SLC1A2 protein lacking only 9 amino acids raises the possibility that wild-type *SLC1A2* might also be oncogenic. We tested this possibility by silencing wild-type *SLC1A2* in AGS cells which are fusion negative. We designed customized *SLC1A2* siRNA specifically targeting exon 1 of *SLC1A2* and also a non-overlapping wild type *SLC1A2* siRNA. As shown in the Figure 3.10, silencing of wild-type *SLC1A2* in AGS cells which are fusion negative with both siRNAs resulted in similar phenotypic effects comparable to *CD44-SLC1A2* silencing in SNU16 cells (Fig. 3.10). In this regard, *CD44-SLC1A2* may be similar to oncogenic fusion genes such as *IgH-Myc* and *TMPPRS2-ERG*, whereby an essentially full-length pro-oncogenic protein is placed under the control of a strong transcriptional promoter.

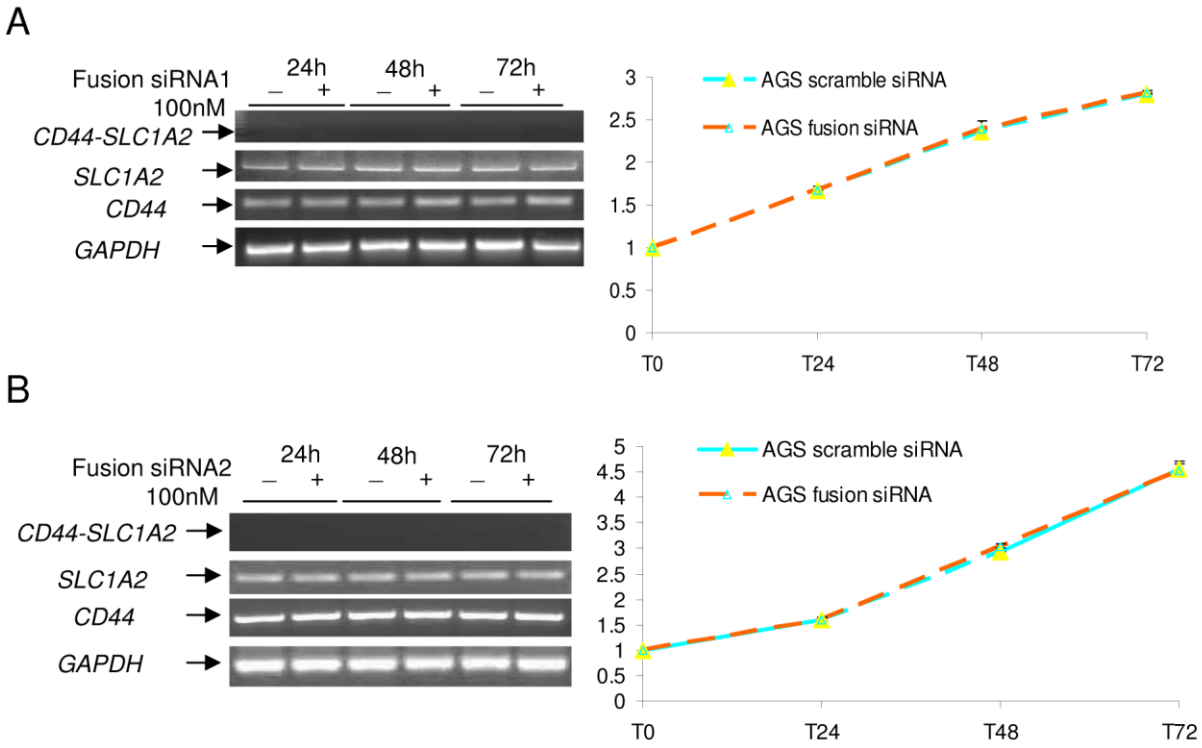


Figure 3.9. *CD44-SLC1A2* silencing does not affect fusion negative AGS cells

A. *Left*, expression of *CD44-SLC1A2*, *SLC1A2* and *CD44* after *CD44-SLC1A2* siRNA1 treatment in AGS fusion negative cells. *CD44*: wild-type *CD44* expression. *CD44* primers were designed to target exons 3-5. *SLC1A2*: wild-type *SLC1A2* expression. *SLC1A2* primers were designed to target exon 1. *GADPH* was used as a loading control. *Right*, proliferation of AGS cells was not affected by fusion-specific siRNA1 silencing.

B. *Left*, expression of *CD44-SLC1A2*, *SLC1A2* and *CD44* after *CD44-SLC1A2* siRNA2 treatment in AGS fusion negative cells. *CD44* primers were designed to target exons 3-5. *SLC1A2*: wild-type *SLC1A2* expression. *SLC1A2* primers were designed to target exon 1. *GADPH* was used as a loading control. *Right*, proliferation of AGS cells was not affected by fusion-specific siRNA2 silencing.

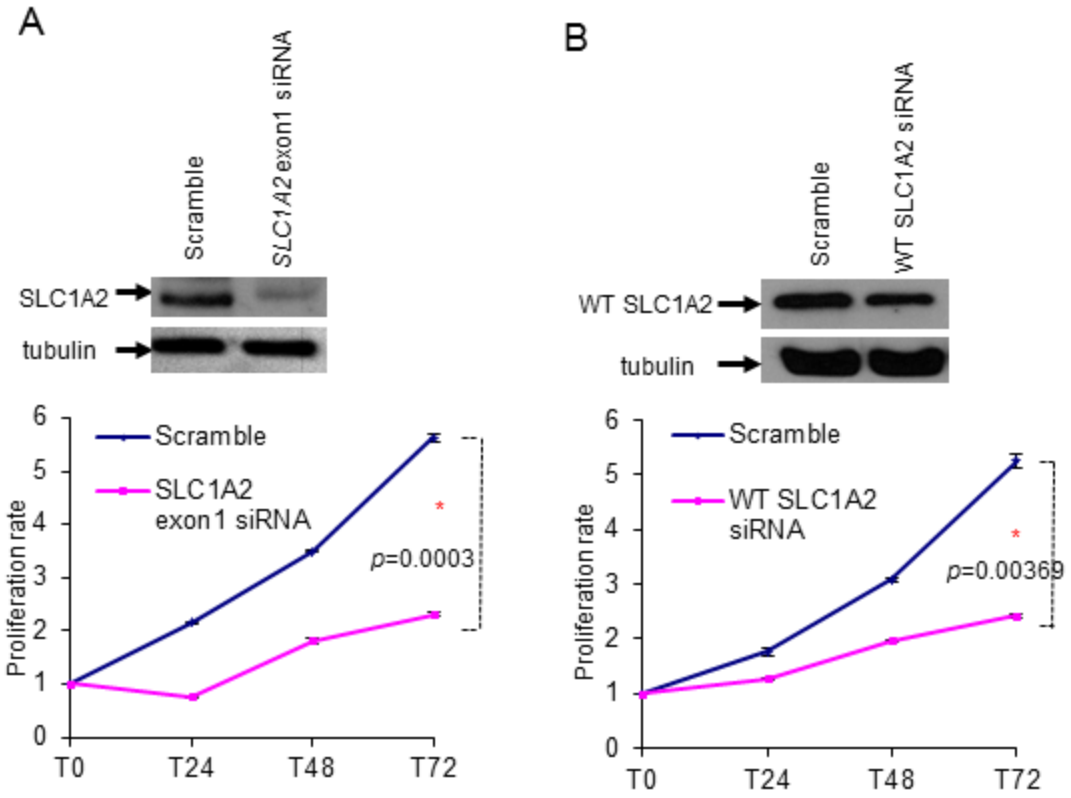


Figure 3.10. Reduction of cellular proliferation in fusion-negative AGS cells after silencing of wild-type *SLC1A2*

A. Silencing of wild-type *SLC1A2* in AGS cells using a siRNA targeting exon 1 of *SLC1A2*. Upper, reduction of SLC1A2 protein expression after *SLC1A2* siRNA treatment. Tubulin control confirmed equal amount of protein loading in both lanes. Lower, AGS cell proliferation after treatment with a scrambled siRNA control or *SLC1A2* exon 1 siRNA. P value was calculated at T72 using student's unpaired t-test.

B. AGS treated with a wild type *SLC1A2* siRNA. Upper, reduction of SLC1A2 protein expression after *SLC1A2* siRNA treatment. Lower, AGS cell proliferation after treatment with a scrambled siRNA control or *SLC1A2* siRNA. P value was calculated at T72 using student's unpaired t-test.

3.3.7 Overexpression of *CD44-SLC1A2* to HFE145 cells

To determine if *CD44-SLC1A2* expression is sufficient to enhance various pro-oncogenic traits, we stably overexpressed *CD44-SLC1A2* in HFE145 normal gastric cells. Compared to control cells, *CD44-SLC1A2* expressing HFE145 cells exhibited enhanced cell proliferation (p=0.007) (Fig. 3.11A). Colony formation assays revealed a signification increase in numbers of colonies formed in cells transfected with *CD44-SLC1A2* vector compared to vector only tranfected cells (p=0.02) (Fig. 3.11B). Overexpression of *CD44-SLC1A2* in HFE145 further resulted in a significant induction of invasion rate compared to vector tranfected cells (*p*=7.75E-05) (Fig. 3.11C). Taken collectively, these results suggest that *CD44-SLC1A2* is likely required by GC cells to maintain several pro-oncogenic traits, such as proliferation, colony formation and invasion.

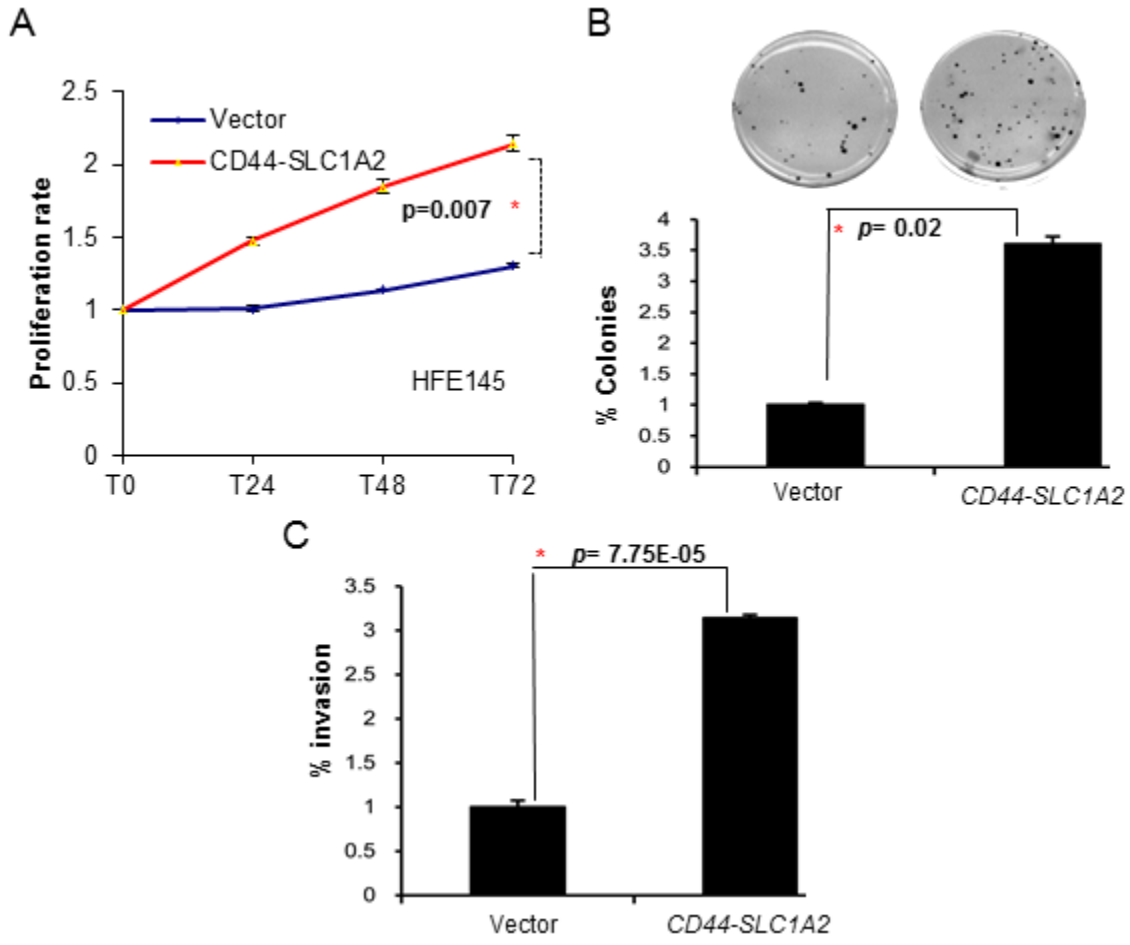


Figure 3.11. Effects of *CD44-SLC1A2* overexpression in HFE145 gastric normal epithelial cells

A. HFE145 cells transfected with *CD44-SLC1A2* expression vector exhibited a significant increase in proliferation rate (p=0.007) compared to vector only tranfected cells. Proliferation rates were measured at 24h, 48h and 72h post tranfection. P value was calculated at T72 using student's unpaired t-test.

B. Colony formation assays revealed a signification increase in numbers of colonies formed in cells transfected with *CD44-SLC1A2* vector compared to vector only tranfected cells (p=0.02). P value was calculated using student's unpaired t-test.

C. Overexpression of *CD44-SLC1A2* in HFE145 also resulted in a significant induction of invasion rate compared to vector tranfected cells (p=7.75E-05). P value was calculated using student's unpaired t-test.

3.3.8 *CD44-SLC1A2* silencing significantly reduces intracellular glutamate levels

One possible mechanism by which *CD44-SLC1A2* may contribute to tumor development is by facilitating glutamate uptake in GC cells. The function of *SLC1A2* as a high-affinity glutamate transporter is intriguing, as previous studies have highlighted an important role for glutamate in cancer metabolism and signaling. Glutamate has been shown to regulate tumor growth and survival (Rzeski et al., 2002; Takano et al., 2001), and gastric tumors have been shown to contain higher glutamate levels compared to normal stomach (Okada et al., 1993). We hypothesized that expression of the *CD44-SLC1A2* gene fusion might act to increase levels of intracellular glutamate in GC. To test this possibility, we measured levels of intracellular glutamate across the GC cell lines. A significantly higher basal glutamate level in *CD44-SLC1A2* expressing SNU16 cells was observed compared to AGS cells (Fig. 3.12, p=0.009). Furthermore, after *CD44-SLC1A2* siRNA treatment, SNU16 glutamate levels were significantly reduced compared to scramble siRNA control (Fig. 3.12, p=0.012). No significant effects were observed in the fusion siRNA treated AGS cells. This observation suggests that *CD44-SLC1A2* may regulate intracellular glutamate levels in GC.

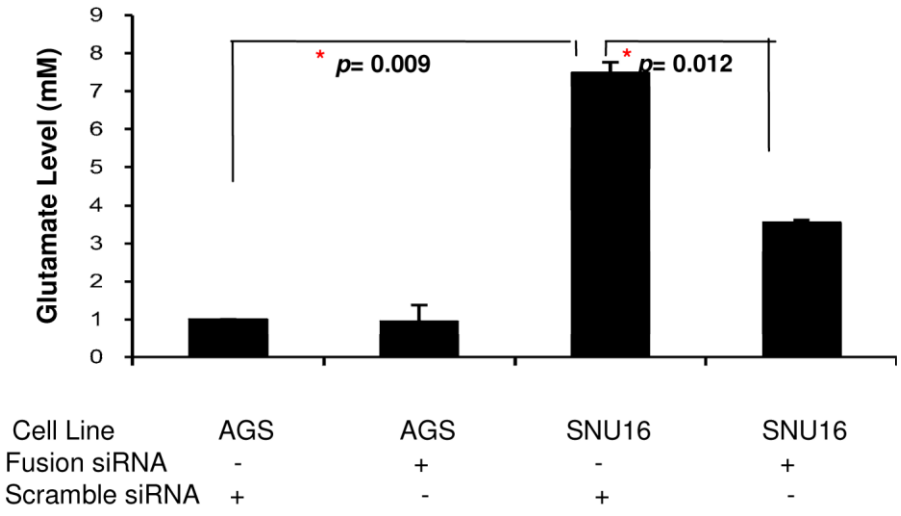


Figure 3.12. *CD44-SLC1A2* regulates intracellular glutamate levels

Glutamate levels in GC cells before and after *CD44-SLC1A2* siRNA treatment. *CD44-SLC1A2* positive SNU16 cells exhibited elevated intracellular glutamate levels compared to fusion negative AGS cells (p=0.009). Knockdown of *CD44-SLC1A2* at T48 resulted in significant decrease of glutamate level in SNU16 cells (p=0.012), but not in AGS cells. All experiments were performed in triplicate. P values were calculated using student's unpaired t-test.

3.3.9 *CD44-SLC1A2* silencing sensitizes GC cells to chemotherapy

Targeting glutamate metabolism in cancer cells has been shown to cause sensitization to pharmacologic treatment (Yang et al., 2009). To test if *CD44-SLC1A2* silencing might sensitize GC cells to drug treatment, we treated control and *CD44-SLC1A2* silenced SNU16 cells to increasing concentrations of cisplatin, a chemotherapy reagent commonly used in GC treatment. GI50s are computed, which is the drug concentration required to cause 50% growth inhibition. We found that SNU16 cells were significantly more sensitive to cisplatin after *CD44-SLC1A2* siRNA treatment (p=1.11x10^{-6}, Fig. 3.13A). The sensitization of *CD44-SLC1A2* silenced cells appears to be specific to cisplatin, as no differences between the control and silenced cells were observed upon treatment with 5-fluorouracil, another GC chemotherapeutic drug (p=0.33, Fig. 3.13B).

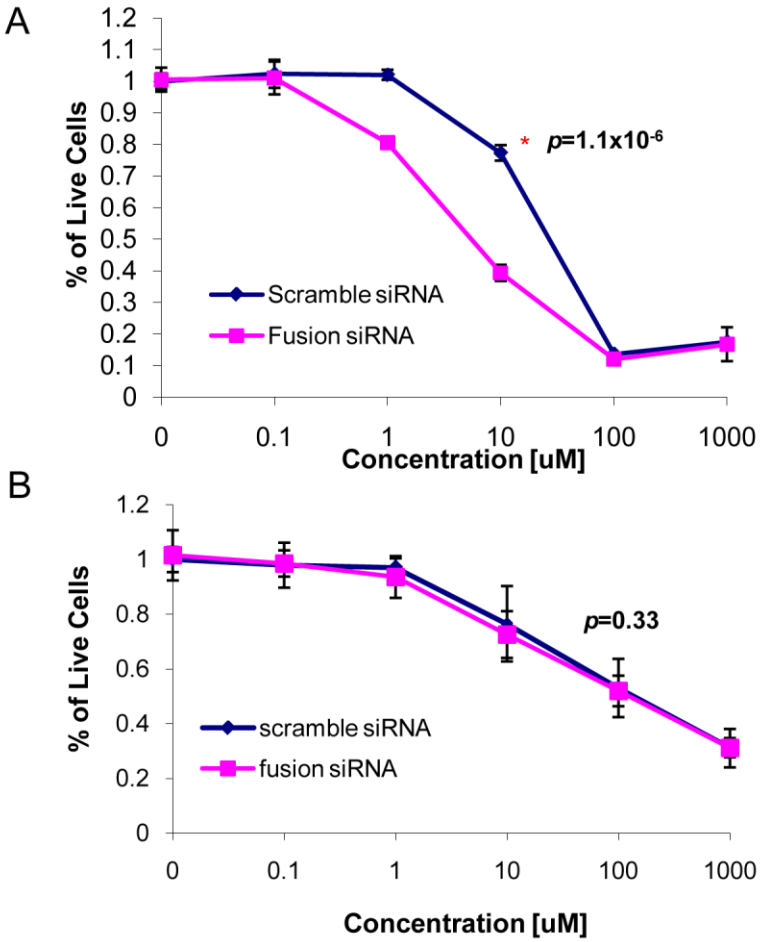


Figure 3.13. ***CD44-SLC1A2*** **sensitizes cells to Cisplatin chemotherapy but not 5-Fluorouracil**

A. Cisplatin sensitivity of SNU16 cells with and without *CD44-SLC1A2* siRNA silencing. SNU16 showed increased sensitivity to cisplatin after *CD44-SLC1A2* siRNA treatment compared to scramble control (p=1.1X10^{-6} at 10uM cisplatin treatment). P value was calculated at Concentration 10uM using student's unpaired t-test.

B. 5-Fluorouracil sensitivity of SNU16 cells before and after *CD44-SLC1A2* siRNA silencing. There was no significant change in sensitivity to 5-Fluorouracil after *CD44-SLC1A2* siRNA treatment compared to treatment with scramble siRNA control (p=0.33). P value was calculated at Concentration 10uM using student's unpaired t-test.

3.4 *CD44-SLC1A2* fusion in primary gastric tumors

3.4.1 Screening of *CD44-SLC1A2* fusion in breakpoint positive samples (Index samples)

To test if *CD44-SLC1A2* might be expressed in clinical specimens, we first assembled RNAs and screened two of the three original index cases exhibiting *SLC1A2* genomic breakpoints (Fig. 3.2B). The third index tumor, GC20021048, had insufficient material available for analysis. Cloning and sequencing of the PCR products confirmed *CD44-SLC1A2* expression in tumor GC2000038, but not in corresponding matched normal tissue (Fig. 3.14A-B). This result demonstrates the *CD44-SLC1A2* expression may occur in primary tumors and it is not a "private" event confined to SNU16 cells alone.

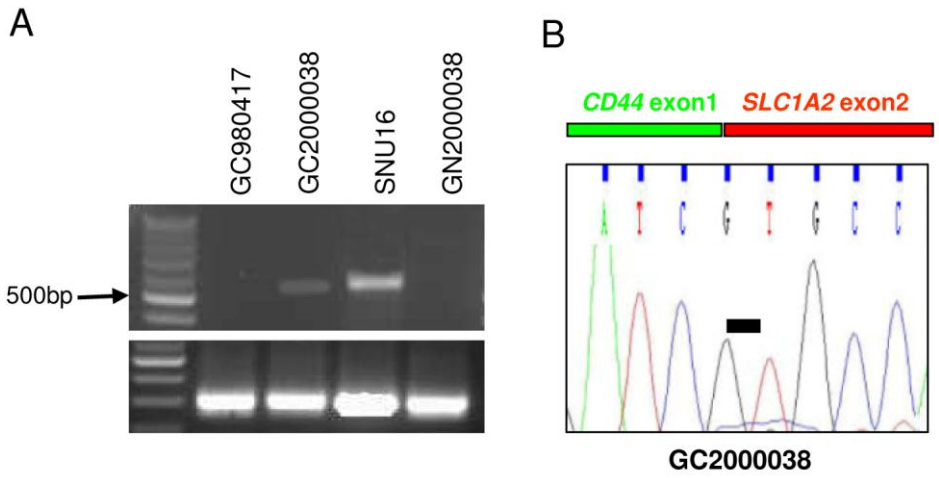


Figure 3.14. *CD44-SLC1A2* expression in index (*SLC1A2* breakpoint positive) primary GCs

A. *CD44-SLC1A2* RT-PCR on two index primary GCs (GC980417 and GC2000038) with *SLC1A2* genomic breakpoints (see Fig. 1). GN2000038 is the matched normal sample to GC2000038. Fusion positive SNU16 was included as a positive control.

B. Sequence of RT-PCR product in (A) revealed the *CD44-SLC1A2* fusion junction in GC2000038. Black bars indicate the fusion junction.

3.4.2 Recurrent *CD44-SLC1A2* fusion in gastric tumor samples

We screened *CD44-SLC1A2* fusion in an independent panel of forty-three gastric tumors and matched gastric normal tissues. We performed *CD44-SLC1A2* RT-PCR using forward primers to *CD44* exon 1 and reverse primers to *SLC1A2* exon 3. Out of 43 tumors tested, two additional tumors amplified a PCR product of similar size to the *CD44-SLC1A2* fusion transcript (Fig. 3.15A), and subsequent cloning and sequencing confirmed the expression of *CD44-SLC1A2* in these tumors (Fig. 3.15B). Similar to the index samples, *CD44-SLC1A2* is not expressed in corresponding matched normal samples (Fig. 3.15A, bottom), supporting the cancer specific nature of this fusion transcript. Subsequent cloning and sequencing of *CD44-SLC1A2* in the fusion-positive tumors revealed that the fusion consistently involved the juxtaposition of *CD44* exon 1 to *SLC1A2* exon 2 (Fig. 3.15B). To rule out the possibility of contamination from SNU16, we further sequenced downstream of fusion transcript, and a nucleotide variant was identified in GC980390 which resulted in a silent mutation (Fig. 3.15C). This apparent requirement for precise joining may be because amongst the *SLC1A2* exons, only exon 2 possesses a suitable alternative start ATG to initiate translation of a near complete SLC1A2 protein.

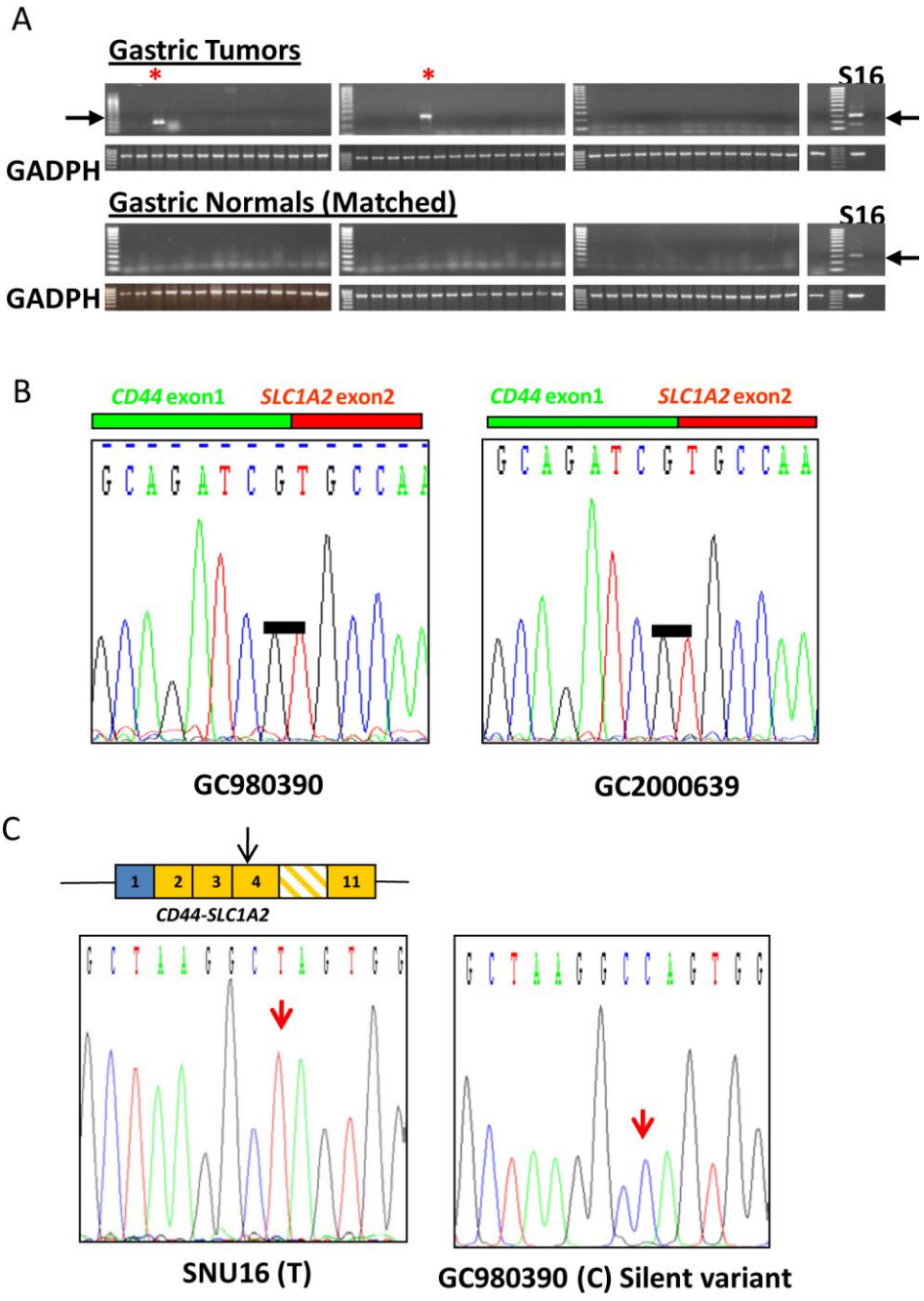


Figure 3.15. *CD44-SLC1A2* expression in large cohort (unselected) of primary gastric tumors and their matched normals

A. *CD44-SLC1A2* RT-PCR on 43 gastric tumors and matched normal gastric tissues. *Top*, gastric tumor samples. Red stars highlight *CD44-SLC1A2* expressing tumors (GC980390, GC2000639). *Bottom*, matched normals. SNU16 was included as a positive control.

B. Cloning and sequencing of *CD44/SLC1A2* RT-PCR products from two fusion positive tumors (GC980390 and GC2000639) confirmed fusion between *CD44* exon 1 and *SLC1A2* exon 2. Black bars indicate the fusion junction.

C. A silent mutation was found in *CD44/SLC1A2* fusion transcript in GC980390. T*op, CD44-SLC1A2* expression construct. Arrow–mutation site. *Bottom*, sequencing of *CD44/SLC1A2* RT-PCR products identified a (T-C) nucleotide change in GC980390. Arrows-nucleotide variants.

3.4.3 Confirmation of *CD44/SLC1A2* genomic inversions in fusion positive primary gastric tumors

In our previous results (Fig. 3.5B), we used long-range PCR with genomic DNA, followed by end-sequencing of the PCR products, to confirm the presence of an intrachromosomal inversion in SNU16. To demonstrate that the genomic inversions may be one of the causal events leading to *CD44/SLC1A2* fusion in primary tumors, we performed long-range PCR using the primers specifically directed against *CD44* exon 1 and the *SLC1A2* intron 1, the same primer pairs used as in SNU16. The results showed the presence of *CD44/SLC1A2* genomic inversions at the DNA level in the two fusion-positive clinical specimens (GC980390 and GC2000038) (Fig. 3.16). Importantly, no such genomic inversion products were observed in the matched normal gastric samples, indicating that the *CD44/SLC1A2* inversion is a cancer-associated somatic event.

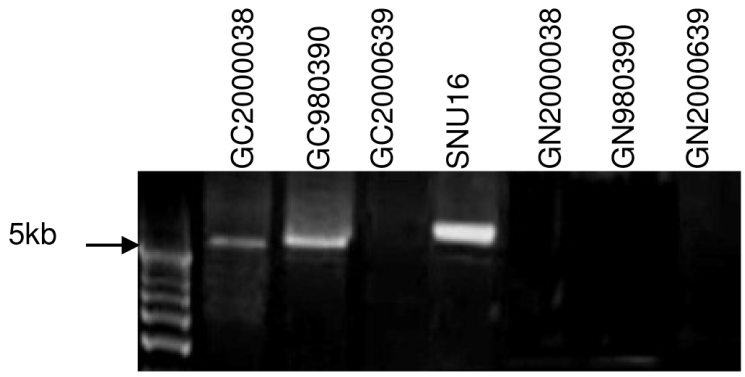


Figure 3.16. Long-range genomic PCR analysis in fusion positive gastric tumor tissues

Long-range PCR confirmed a chromosomal inversion of *CD44-SLC1A2* in two fusion positive gastric tumors (GC2000038 and GC980390). GN2000038, GN980390 and GN2000639 are matched gastric normal controls. SNU16 was used as a positive control. Primers used were the same as Figure 3.5B.

3.4.4 Tumors expressing high *SLC1A2* levels are associated with *CD44-SLC1A2* positivity

The *CD44* gene is both activated by Wnt signaling and repressed by p53, two common cancer-related pathways. As such, *CD44* is highly expressed in many cancers including GC (Kim et al., 2009). Therefore, one consequence of the *CD44-SLC1A2* fusion might be to place *SLC1A2* under the regulatory control of *CD44* promoter elements, causing high levels of *SLC1A2* expression in tumors. This model predicts that tumors expressing high *SLC1A2* levels may also tend to be *CD44-SLC1A2* positive. To explore this possibility, we queried a previously described gene expression database of 197 gastric cancers to identify tumors expressing high *SLC1A2* levels (Ooi et al., 2009). We screened fifteen GCs from the top 15 percent of *SLC1A2* overexpressing tumors for *CD44-SLC1A2* expression. Among the fifteen tumors, five GCs expressed the *CD44-SLC1A2* fusion transcript (Fig. 3.17, blue crosses), and none of the matched adjacent normal tissues expressed *CD44-SLC1A2* (Fig. 3.17). Thus, while the rate of *CD44-SLC1A2* positivity in an unselected patient cohort is low (1-2%), the *CD44-SLC1A2* positivity rate is elevated in this selected subpopulation (33%, 5 out of 15 tumors). This result is consistent with the *CD44-SLC1A2* fusion causing the transcriptional upregulation of *SLC1A2*. In *CD44-SLC1A2* negative tumors, high *SLC1A2* levels may be attributed to alternative mechanisms, such as focal genomic amplification, fusion to other partners, and EGF or mTOR/Akt signaling (Wu et al., 2010; Zelenaia et al., 2000).

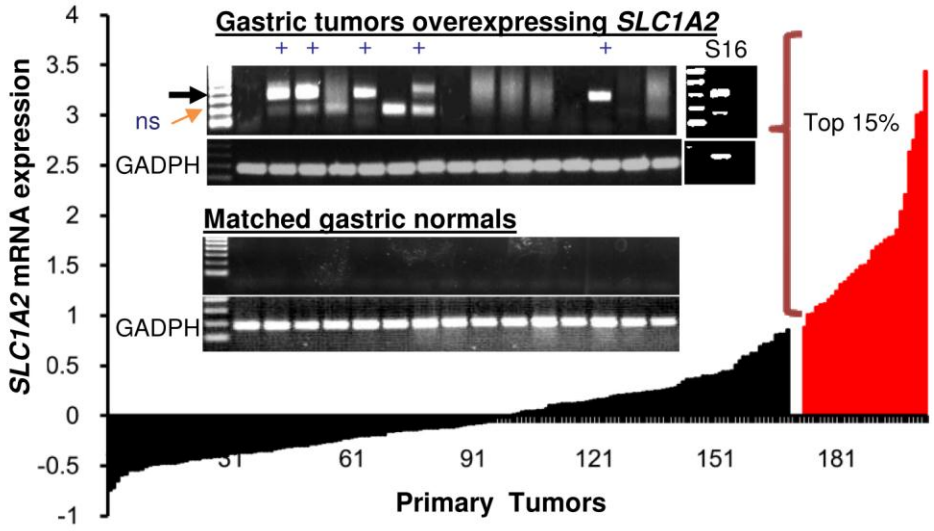


Figure 3.17. *CD44-SLC1A2* positive tumors are associated with high *SLC1A2* expression

Graph, x-axis - 197 GCs sorted by levels of *SLC1A2* expression. Y-axis: Log transformed gene expression data were median-centered. The top 15% of high *SLC1A2* expressing tumors are shown in red. *Inset*, RT-PCR screening of *CD44-SLC1A2* in the top 15% of high *SLC1A2*-expressing GCs. *GADPH* was used as a loading control. Samples expressing *CD44-SLC1A2* are highlighted using blue crosses. SNU16 cells (S16) were included as a positive control. The smaller band of 200bp was sequenced and identified to be non-specific (ns). *CD44-SLC1A2* RT-PCR analysis of the matching 15 normal gastric tissues was also shown.

3.4.5 Glutamate levels in primary gastric tumors

In many cancers, glutamate and its related amino acid glutamine function as important amino acids regulating tumor growth and survival (Rzeski et al., 2002; Takano et al., 2001), and gastric tumors have been shown to contain higher glutamate levels compared to normal stomach tissues (Okada et al., 1993). Previously, we have shown that *CD44-SLC1A2* silencing significantly reduces intracellular glutamate levels in gastric cancer cells (Fig. 3.11). Next, to assess the levels of glutamate in our in-house primary GCs, we used the glutamate assay to measure glutamate levels in 20 pairs of matched tumor and normal pairs. Significantly elevated levels of glutamate were detected in primary tumors compared to matched normal stomach controls (n=20, p=0.038, paired t-test) (Fig. 3.18), which confirm the results shown in previous studies.

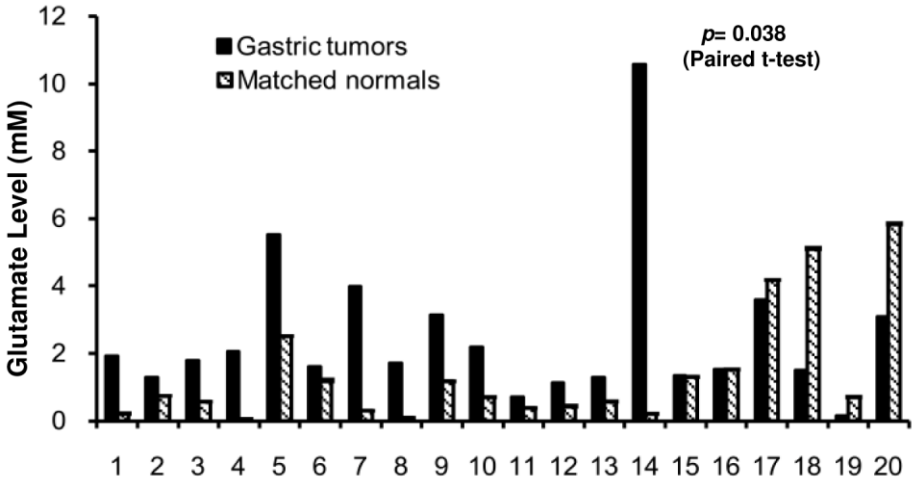


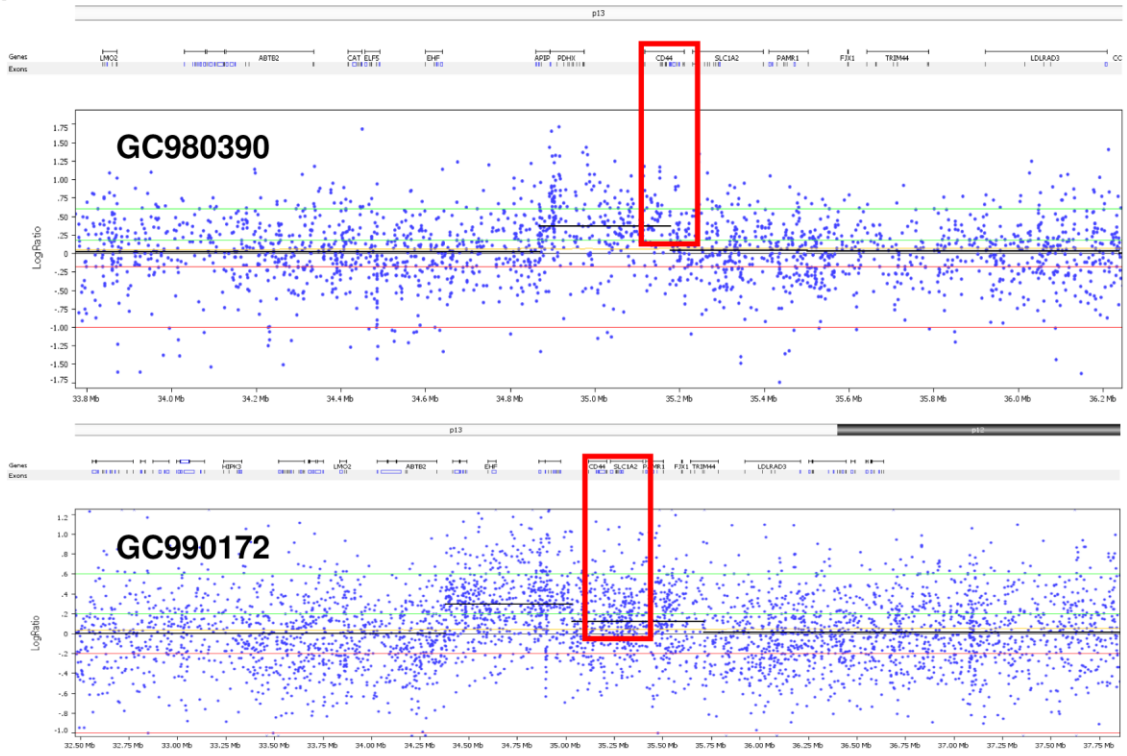
Figure 3.18. Glutamate levels in primary GCs

Glutamate levels were significantly higher in gastric tumor tissues compared to their matched normals (p=0.038). x-axis: 20 cancer/normal pairs (Gastric cancer – black, Matched normal – grey). y-axis: glutamate concentration. P-values were computed using paired t-test.

3.4.6 *CD44-SLC1A2* expression can occur independently of 11p13 amplification

Although *CD44-SLC1A2* was initially identified in tumors exhibiting 11p13 amplification (Fig. 3.2B), 11p13 amplification may not be essential for *CD44-SLC1A2* fusion expression. To investigate the relationship between 11p13 genomic amplification and *CD44-SLC1A2* expression, we analysed seven fusion positive tumors using Affymetrix SNP6 arrays. Of seven fusion-positive tumors, two tumors (GC980390 and GC990172) exhibited evidence of 11p13 genomic amplification, while the other five did not (Fig. 3.19). This finding demonstrates that *CD44-SLC1A2* expression may be observed in tumors independent of 11p13 genomic amplification. Further supporting the notion that 11p13 amplification and *CD44-SLC1A2* gene fusion are distinct events, a comparison of *CD44* and *SLC1A2* expression levels across 45 gastric tumors, including a) 11p13 non-amplified samples (32 samples); b) 11p13 amplified but fusion-negative samples (6 samples), and c) *CD44-SLC1A2* fusion positive samples (7 samples) revealed that high *SLC1A2* expression levels may be more strongly associated with fusion-events rather than generalized 11p13 amplification (Fig.3.20).

A


GC980390
LogRatio
GC990172
LogRatio

B

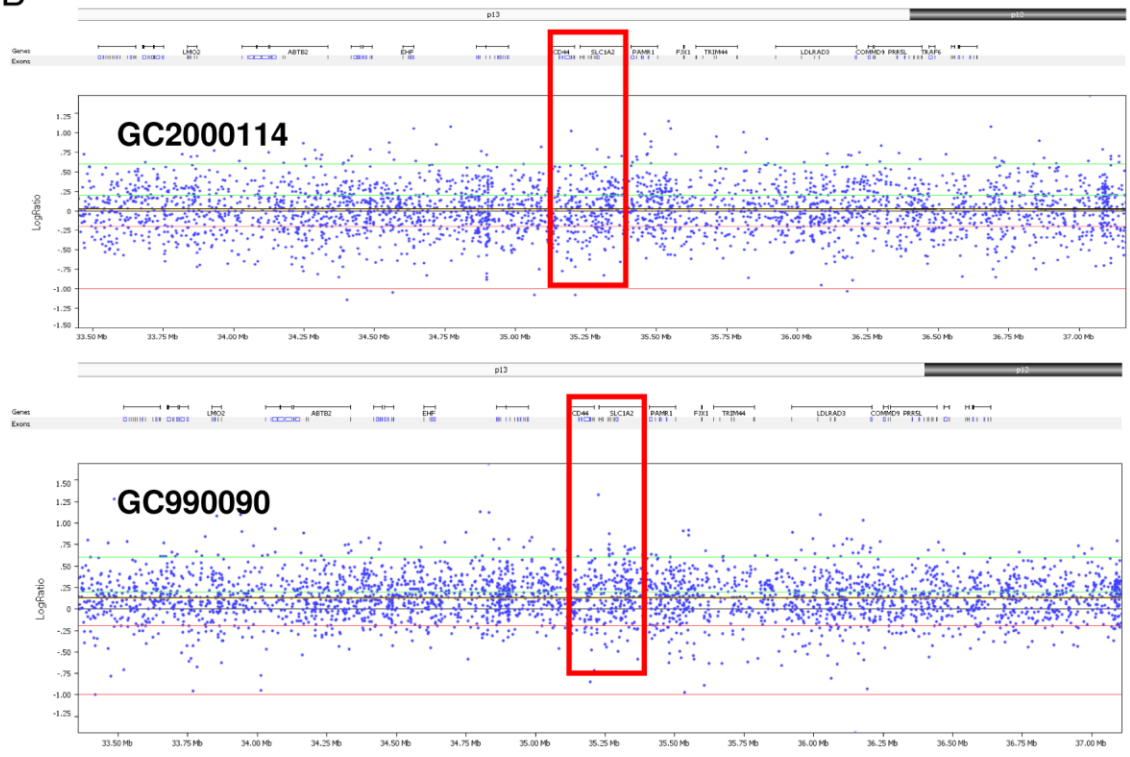

GC2000114
LogRatio
GC990090
LogRatio

Figure 3.19. 11p13 copy number status in *CD44-SLC1A2* expressing samples

Seven *CD44-SLC1A2* expressing tumors were analyzed for 11p13 copy number amplification using Affymetrix SNP6 arrays. The X-axis indicates genomic position and the y-axis is the log transformed copy number data. Each blue dot represents a SNP array probe. Segmented copy number data are plotted as horizontal black lines. The *CD44* and *SLC1A2* gene region is highlighted as a red rectangle

(A) Two fusion-positive samples (GC980390 and GC990172) exhibiting 11p13 amplification.

(B) Two fusion-positive samples (GC2000114 and GC990090) without 11p13 amplification.

3.4.7 Impact of 11p13 amplifications and *CD44-SLC1A2* fusions on *SLC1A2* and *CD44* expression

We compared *CD44* and *SLC1A2* expression levels across 45 gastric tumors, including a) 11p13 non-amplified samples (32 samples); b) 11p13 amplified but fusion-negative samples (6 samples), and c) *CD44-SLC1A2* fusion positive samples (7 samples). The rate of 11p13 amplification in this series (~17%) is similar to frequencies previously reported in the literature (Fukuda et al., 2000). It is important to note that in this experiment, the expression measurements were inferred using U133P2 Affymetrix microarray probes, which target the 3' ends of genes. Compared to non-amplified samples, fusion positive samples exhibited significantly increased 3' *SLC1A2* gene expression (p=0.004), but 11p13 amplified samples did not (p=0.86) (Fig. 3.20A). These findings suggest that high *SLC1A2* expression levels may be driven more by fusion-events rather than generalized 11p13 amplification The one exception was a sample with a high-level focal 11p13 amplification (GC980417) – in this tumor, *SLC1A2* was highly expressed (Fig. 3.20A). Intriguingly, unlike *SLC1A2*, a very different scenario was observed for *CD44*. Specifically, while *CD44* 3' transcripts were significantly overexpressed in 11p13 amplified tumors (p=0.016), they were significantly underexpressed in fusion positive tumors (p=0.006) (Fig. 3.20B). We postulate that this latter finding may be due to the *CD44/SLC1A2* genomic inversion decoupling the 3' end of the *CD44* gene (the region detected by the Affymetrix array) from the endogenous *CD44* promoter. Additional evidence of this decoupling was obtained in a real-time PCR analysis measuring *SLC1A2* exon 1, where unlike the 3' *SLC1A2* transcripts, *SLC1A2* exon 1 (which is not part of the

CD44-SLC1A2 fusion) was not observed to be highly expressed relative to non-amplified samples in fusion positive samples (Fig. 3.20C-D).

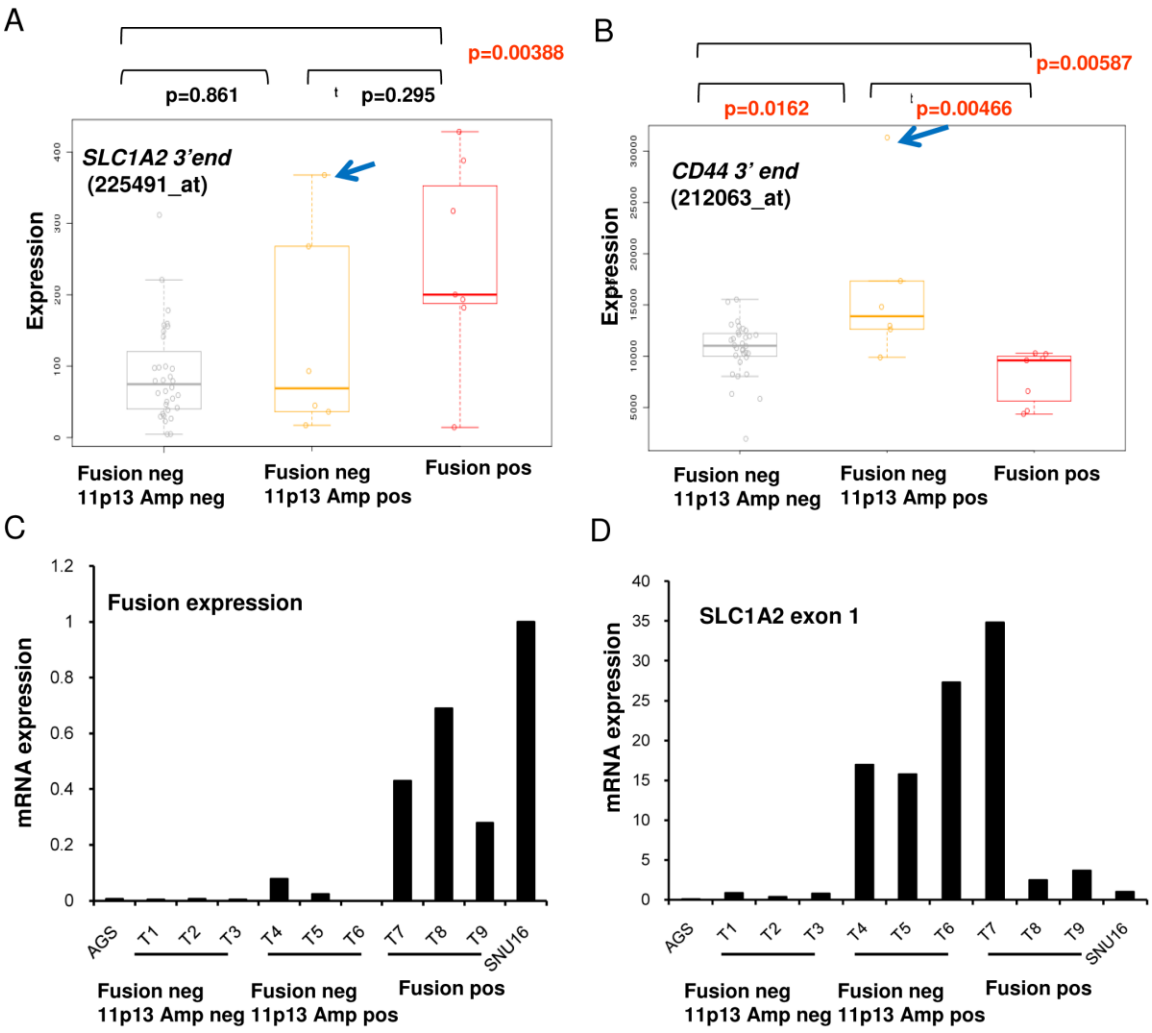


Figure 3.20. *CD44* and *SLC1A2* expression levels of 11p13 non-amplified, 11p13 amplified, and fusion positive samples

A-B. *CD44* and *SLC1A2* expression in 45 gastric tumor samples profiled on both gene expression and SNP microarrays. Of these 45, 32 samples are 11p13 non-amplified (blue), 6 samples are 11p13 amplified but fusion-negative (yellow), and 7 samples are *CD44-SLC1A2* fusion positive (red). Significant p-values ($p<0.05$, Student's t-test) are shown in red. Whiskers represent the range of data. Borders of the box indicate 25th percentile and 75th percentile of the data respectively. The bar inside each box indicate average value of the data in each category.

A. *SLC1A2* gene expression measured using Affymetrix U133 plus probe 225491_at. Fusion positive samples show significantly increased *SLC1A2* gene expression compared

to non-amplified samples (p=0.004), but 11p13 amplified samples do not (p=0.86). The blue arrow highlights one 11p13 amplified sample (GC980417) with a high-level focal amplification in the *CD44/SLC1A2* gene region, suggesting that in some cases 11p13 amplification can drive increased *SLC1A2* gene expression.

B. *CD44* gene expression measured using Affymetrix U133 plus probe 212063_at. 11p13-amplified samples showed significantly increased *CD44* expression compared to non-amplified samples (p=0.016) but fusion positive samples showed a significantly *decreased* level of *CD44* expression (p=0.0059), consistent with the *CD44/SLC1A2* inversion event decoupling the wild-type *CD44* gene from its endogenous promoter.

C-D. Real-time PCR analysis of *CD44-SLC1A2* expression (targeting the fusion junction), and *SLC1A2* (targeting *SLC1A2* exon 1). The tumors were stratified into 3 groups. Group 1: T1-3, 11p13 non-amplified and fusion negative; Group 2: T4-6, 11p13 amplified and fusion negative; and Group 3: T7-9, 11p13 non-amplified but fusion positive. AGS (fusion-negative) and SNU16 (11p13 amplified, fusion positive) were also included. All readings were normalized against a *GADPH* housekeeping control. Each data point represents the average of two independent experiments.

C. *CD44-SLC1A2* fusion expression. Only Group 3 fusion-positive tumors and SNU16 cells were observed to express *CD44-SLC1A2*.

D. *SLC1A2* exon 1 expression. 11p13-amplified samples were observed to express high levels of *SLC1A2* exon 1, consistent with these samples exhibiting "copy number driven" expression. In contrast, *SLC1A2* exon 1 was not highly expressed in 2 out of 3 fusion positive samples, consistent with the high levels of 3' *SLC1A2* transcript in these samples being driven by fusion to *CD44*.

3.4.8 Unsupervised clustering of GC expression profiles reveals clustering of *SLC1A2*-high expressing tumors

Next, we explored if *SLC1A2*-high expressing tumors might be associated with any distinct clinicopathogic traits. To study whether the *SLC1A2*-high expressing tumors exhibit molecular similarity and whether the fusion subclass emerges as a discrete subset, we performed an unsupervised clustering analysis of the 197 gastric tumors. The results revealed that a subclass of high *SLC1A2*-expressing tumors tended to cluster together, suggesting that high *SLC1A2* expression may define a distinct molecular subgroup of GC (Fig. 3.21). To identify predominant biological themes associated with this molecular subgroup, we performed gene ontology analysis on a 710-gene '*SLC1A2* signature', generated by comparing the top 15% of *SLC1A2* high-expressing tumors against the bottom 15% (Wilcoxon signed ranked test, FDR =0.005). Genes expressed in *SLC1A2* high expressing tumors were associated with ribosomal biosynthesis and protein translation (corrected $p=5.12x10^{-33}$; Fisher test, Table 3.2). These results suggest that a subgroup of tumors expressing high *SLC1A2* levels, either by *CD44* fusion or alternative mechanisms, may comprise a distinct subclass of GC.

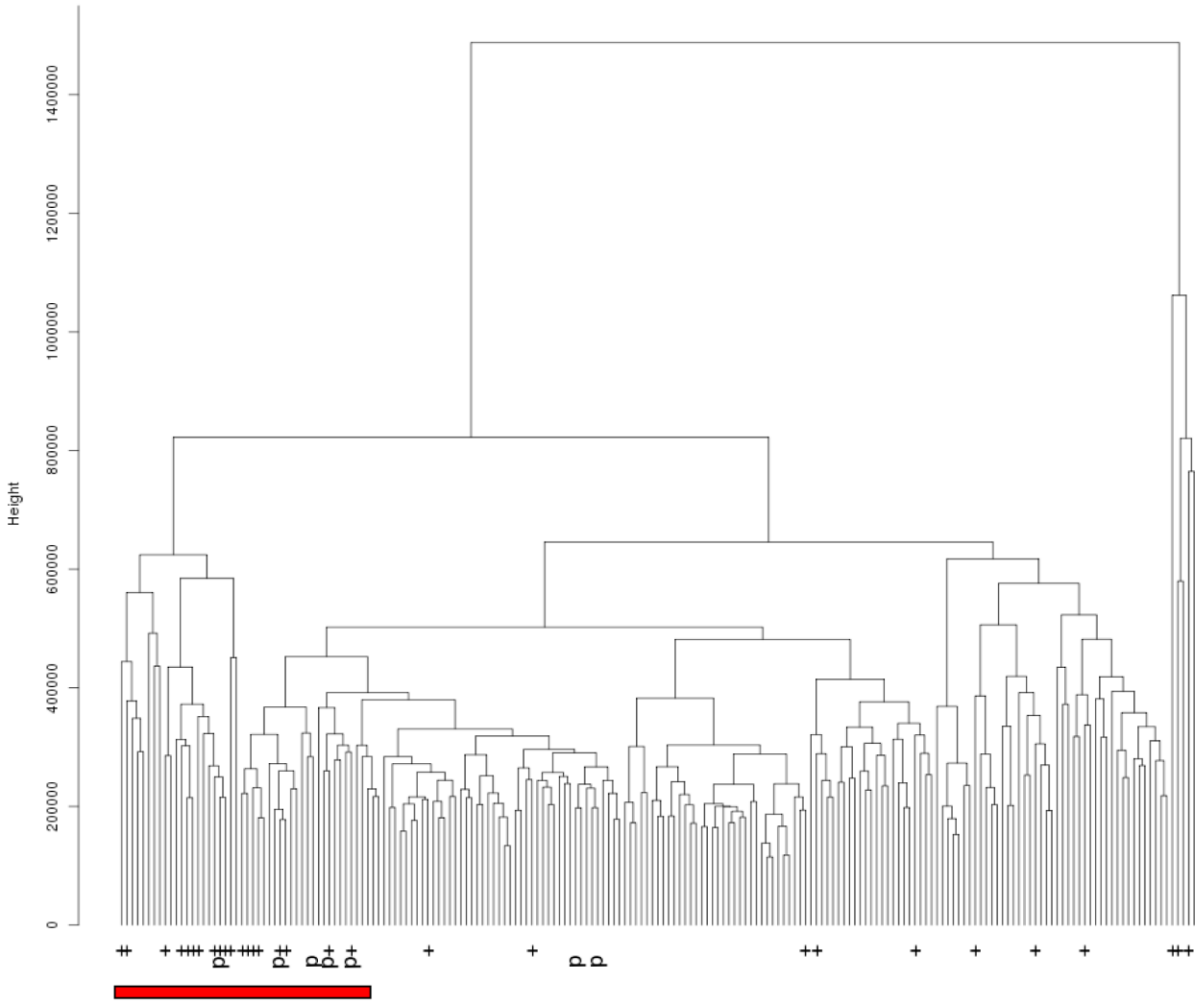


Figure 3.21. Unsupervised clustering of GC expression profiles reveals clustering of *SLC1A2*-high expressing tumors

197 GC gene expression profiles were clustered using an unsupervised hierarchical clustering algorithm. Samples with high *SLC1A2* expression (top 15%) were indicated with a '+'. A subgroup of the high *SLC1A2* expressing tumors were observed to cluster close to one another (red bar).

Table 3.2 : Gene ontology analysis of *SLC1A2*-high expressing tumors

Gene Ontologies upregulated in *SLC1A2*-high tumors

Category	Term	Count	%	PValue	Benjamini
GOTERM_BP_FAT	GO:0006414~translational elongation	39	10.99	3.94E-39	5.79E-36
SP_PIR_KEYWORDS	Ribosome	32	9.01	1.43E-35	5.12E-33
KEGG_PATHWAY	hsa03010:Ribosome	34	9.58	3.80E-33	4.41E-31
GOTERM_CC_FAT	GO:0022626~cytosolic ribosome	32	9.01	4.83E-32	1.24E-29
GOTERM_MF_FAT	GO:0003735~structural constituent of ribosome	37	10.42	1.97E-28	9.38E-26
SP_PIR_KEYWORDS	ribosomal protein	38	10.70	2.29E-28	4.11E-26
GOTERM_CC_FAT	GO:0033279~ribosomal subunit	33	9.30	2.41E-26	3.08E-24
GOTERM_CC_FAT	GO:0005840~ribosome	39	10.99	3.07E-25	2.62E-23
GOTERM_CC_FAT	GO:0044445~cytosolic part	34	9.58	5.23E-25	3.35E-23
SP_PIR_KEYWORDS	ribonucleoprotein	40	11.27	4.39E-24	5.26E-22

Gene Ontologies downregulated in *SLC1A2*-high tumors

Category	Term	Count	%	PValue	Benjamini
SP_PIR_KEYWORDS	Acetylation	49	39.84	1.19E-12	2.76E-10
SP_PIR_KEYWORDS	phosphoprotein	75	60.98	1.45E-07	1.67E-05
GOTERM_CC_FAT	GO:0043233~organelle lumen	28	22.76	7.16E-05	8.66E-03
GOTERM_CC_FAT	GO:0031974~membrane-enclosed lumen	29	23.58	3.65E-05	8.84E-03
GOTERM_CC_FAT	GO:0070013~intracellular organelle lumen	27	21.95	1.32E-04	1.06E-02
GOTERM_CC_FAT	GO:0031981~nuclear lumen	23	18.70	3.01E-04	1.81E-02
UP_SEQ_FEATURE	short sequence motif:Nuclear localization signal	11	8.94	4.20E-05	2.43E-02
SP_PIR_KEYWORDS	repressor	11	8.94	4.28E-04	3.24E-02

Category – Database; Term - Enriched ontology term; Count - Number of genes overlapping with the *SLC1A2* signature; % - percentage of overlapping genes from input signature list; P-value - Modified Fisher exact test pvalue; Benjamini - corrected p value using the Benjamini method.

Chapter IV: Discussion

The cancer-specific nature of fusion genes have earned them an important place in many translational cancer research applications, including molecular subtyping, monitoring for disease relapse, and as drug targets. In pediatric ALL, expression of *AML-ETO* and *PML-RAR* are routinely used to diagnose particular clinical subtypes (Ghaffari et al., 2006). The treatment of CML has been revolutionized by imatinib, an inhibitor of the *BCR-ABL* fusion gene (Deininger et al., 2005). Along with *AGTRAP-BRAF* fusions (Palanisamy et al., 2010), *CD44-SLC1A2* represents another recurrent gene fusion identified in a major GI cancer, providing further evidence for the existence of this important class of molecular aberrations in GI malignancies.

In this study, we used Genomic Breakpoint Analysis (GBA) to uncover the existence of *CD44-SLC1A2* gene fusions in GC. Although our starting microarray data set of GC CNAs is similar to data sets used in regular array-based Comparative Genomic Hybridization (aCGH) studies (Niini et al., 2010), our analytical strategy differs fundamentally from aCGH. Specifically, in conventional aCGH, attention is primarily focused on genes residing in the central regions of amplicons, while in our approach, amplicon boundaries and edges assume greater relevance. Notably, while GBA has been previously used for fusion gene discovery in leukemia (Kawamata et al., 2008; Mullighan et al., 2007), our study demonstrates that this approach can also highlight potential fusion genes in solid epithelial tumors.

Amongst genes exhibiting genomic breakpoints, we prioritized genes for study based on their rates of recurrence in multiple samples and the specificity of the breakpoints. Using these two criteria, we noted that only a minor portion of candidate genes exhibited recurrent breakpoints (N>=2, Table 3.1). In addition, for the majority of genes exhibiting genomic breakpoints in multiple samples (eg *CRKRS*, *TTC25*), the breakpoints were randomly scattered throughout the gene body consistent with a random breakage model of chromosomal amplification. *SLC1A2* was nominated for further experimental studies as four GCs out of 133 exhibited genomic breakpoints specifically localized to the 5' region of the *SLC1A2* gene (Fig. 3.2 B).

GBA does come with a few caveats, as fusion events arising from balanced chromosomal rearragements would not alter overall copy number levels and are unlikely to be detected. Whether the fusion genes identified by breakpoint analysis are similar or different to those identified by other methods such as COPA, transcriptome sequencing, and genomic paired-end sequencing is yet unclear (Bignell et al., 2007; Rubin and Chinnaiyan, 2006). Applying these distinct discovery methods on a common set of cancer samples would prove useful to clarify the specific pros and cons of each approach. However, GBA has the advantage of being readily applicable to array-CGH data, for which there are numerous large scale data sets available in the public domain (Beroukhim et al., 2010). Revisiting these data sets may identify additional genes recurrently targeted by genomic breakpoints in solid cancers.

Chromosomal amplification events are often associated with complex patterns of local genome rearrangement (Bignell et al., 2007), and the *CD44-SLC1A2* fusions were identified by tumors and cell line exhibiting *SLC1A2* genomic breakpoints in regions of generalized 11p13 amplification (Fig. 3.2B). SKY and FISH analysis confirmed chromosomal rearrangement associated with 11p13 amplification in SNU16 (Fig. 3.3). Given the close genomic proximity of *CD44* and *SLC1A2*, it is likely that the *CD44-SLC1A2* fusion is caused by a paracentric inversion resulting in the physical fusion of the *CD44* upstream region to the *SLC1A2* gene (Fig. 3.4C). Supporting this model, genome rearrangements and inversions within the *CD44/SLC1A2* region in SNU16 cells and fusion positive tumors were demonstrated using fiber-FISH and high resolution long-range PCR (Fig. 3.5, Fig. 3.16). Emerging evidence indicates that genes in close proximity may represent favorable targets for gene fusion events in cancer. For example, in prostate cancer the tendency of *TMPRSS2* to fuse with *ERG* rather than *ETV1* or *ETV4*, when all three genes encode ETS transcription factors, is likely due to *TMPRSS2* lying only 3 Mb away from *ERG* on Chr 21, with both genes being brought together in close proximity due to androgen receptor mediated interactions (Lin et al., 2009; Mani et al., 2009). Likewise, in breast cancer, production of the *BCAS3/BCAS4* fusion gene is also likely facilitated by local genomic rearrangements affecting both genes on Chromosome 17 (Bärlund et al., 2002).

The identification of *SLC1A2*, a glutamate transporter, as a fusion gene participant is notable. To date, the vast majority of known oncogenic fusion events have largely involved transcription factors (eg *Myc*, *RAR*, etc) or kinases (eg *BCR-ABL*) (de Klein et

al., 1982; Edwards, 2010; Ghaffari et al., 2006; Rabbitts and Boehm, 1991). The discovery of *CD44-SLC1A2* raises the intriguing possibility that oncogenic gene fusions may also target genes involved in cancer metabolism. Importantly, the notion that glutamate is relevant for cancer development is not without precedent. Indeed, a substantial body of evidence has implicated glutamate and glutamine as a critical amino acid necessary for the maintenance and elaboration of many cancer-specific traits (DeBerardinis and Cheng, 2010; Medina et al., 1992). For example, glutamate and glutamine have been shown to regulate tumor growth and control oncogenic signals such as mTOR (Nicklin et al., 2009; Rzeski et al., 2002; Takano et al., 2001). The requirement of cancer cells for glutamate may also be related to the Warburg effect - a universal feature of cancer cells where they exhibit overactive glycolysis due to a deficiency in channeling glycolytic metabolites into the TCA cycle for ATP generation. Besides glycolysis, glutamate may provide cancer cells with an alternative route of ATP production since intracellular glutamate and glutamine can also be converted into alpha-ketoglutarate, a TCA cycle intermediate (Dang, 2010).

Our study also suggests that *CD44-SLC1A2* may function to facilitate glutamate accumulation in GC cells. The *CD44-SLC1A2* gene fusion is predicted to produce a slightly truncated *SLC1A2* protein that retains most of the key protein domains required for glutamate transporter fusion, and protein immunoflouresence studies revealed high *SLC1A2* immunoreactivity at the cell membrane in SNU16 cells, consistent with its role as a glutamate transporter (Fig. 3.6). Moreover, silencing of *CD44-SLC1A2* reduced levels of intracellular glutamate (Fig. 3.11), and is associated with the inhibition of

several pro-oncogenic phenotypes such as cellular proliferation, invasion and colony formation (Fig. 3.7, Fig. 3.8). Glutamate levels have been shown to be elevated in many cancers, and in our study we confirmed that glutamate levels are also elevated in gastric tumors compared to normal stomach, consistent with a previous report (Okada et al., 1993) (Fig. 3.18).

Our ability to study the effects of endogenous *CD44-SLC1A2* inhibition was limited to SNU16 cells since this is the only fusion-positive GCCL. Therefore, we have further demonstrated a pro-oncogenic role for *CD44-SLC1A2* in two ways. First, because *CD44-SLC1A2* is predicted to produce a slightly truncated but almost full-length SLC1A2 protein, it is possible that wild-type SLC1A2 should exhibit pro-oncogenic effects. We tested this possibility by silencing wild-type *SLC1A2* in AGS cells which is fusion negative. As shown in Figure 3.10, we observed significant phenotypic effects in the AGS cells, similar to *CD44-SLC1A2* knockdown in SNU16 fusion-positive cells. These results suggest a pro-oncogenic role for both *CD44-SLC1A2* and *SLC1A2*. In this regard, *CD44-SLC1A2* is similar to other recognized fusion genes such as *IgH-Myc* and *TMPPRS2-ERG*, where an essentially wild-type pro-oncogenic protein is placed under the control of a strong transcriptional promoter. Second, we ectopically expressed *CD44-SLC1A2* in HFE145 gastric epithelial cells (Fig. 3.12). Compared to control cells, *CD44-SLC1A2* overexpressing HFE145 cells demonstrated enhanced cellular proliferation, colony formation, and invasion, once again supporting a pro-oncogenic role for *CD44-SLC1A2* (Fig. 3.12).

Beyond its effects in cancer development, targeting *CD44-SLC1A2* in fusion-positive tumors may also represent a promising avenue for sensitizing GCs to commonly used standard-of-care chemotherapies, since silencing *CD44-SLC1A2* was sufficient to cause a significant sensitization of GC cells to cisplatin *in vitro* (Fig. 3.13). It will be interesting to evaluate the potential of *CD44-SLC1A2* as a drug target, and determining exactly how this gene fusion, and possibly other glutatmate-related transporters, may contribute to GC development by establishing a metabolic environment favouring oncogenesis.

To extend this study from *in vitro* to *in vivo*, we also screened *CD44-SLC1A2* fusion in primary gastric tumors. The initial rate of *CD44-SLC1A2* positivity in an unselected GC patient cohort was low (2/43) (Fig. 3.15), this frequency was markedly increased when analysis was confined to a specific patient sub-population (tumors with high *SLC1A2* expression) (Fig. 3.17). We believe that this result underscores the reality that most solid epithelial cancers are likely to comprise a heterogeneous collection of distinct biological subtypes, each with distinctive patterns of genetic aberrations and oncogenic pathway activity. It is worth noting that the issue of tumor heterogeneity has not been explicitly considered in many previous cancer genome studies (Wood et al., 2007), and our results suggest that this issue should be incorporated in future fusion gene discovery efforts.

The absolute rate of *CD44-SLC1A2* positivity appeared to be relatively low in this study, however analyses of larger GC patient cohorts will be required to determine the true *CD44-SLC1A2* positivity rate. Nevertheless, we note that even low frequency events in cancer can prove therapeutically useful, as shown *EML4-ALK* fusions in lung cancer (1-

5%) (Soda et al., 2007) and *RAF* fusions in gastric, melanoma and prostate cancers (Palanisamy et al., 2010).

We cloned and sequenced the *CD44-SLC1A2* fusion transcript from 6 fusion-positive tumors, including SNU16 cells. In all cases, the fusion involved a juxtaposition of *CD44* exon 1 to *SLC1A2* exon 2. This recurrent juxtaposition is not unexpected, since amongst the *SLC1A2* exons, only *SLC1A2* exon 2 possesses a suitable alternative start ATG to initiate translation of a near complete *SLC1A2* protein. This information further supports the functional significance of the *CD44-SLC1A2* fusion. We identified genetic variations in the *CD44-SLC1A2* sequences from different fusion-positive cases, demonstrating that they are distinct entities. These sequence polymorphisms, occurring in *SLC1A2* exon 4 corresponding to a silent T/C variant, were confirmed by resequencing both strands (Fig. 3.15C). These findings collectively indicate that *CD44-SLC1A2* expression is recurrent.

Most of the *SLC* family gene fusions identified to date have involved the *SLC* gene at the 5' end of the fusion, such as *SLC45A3-BRAF* in prostate cancer and *SLC34A2-ROS1* in non-small-cell lung cancer. To determine if there is a productive 3' partner for *SLC1A2*, we performed 3’ RACE from exon1 of *SLC1A2* in SNU16 cells. However, only wild-type *SLC1A2* products were detected downstream from *SLC1A2* exon 1. In addition, due to limited tissue availability, we could not perform 3’ RACE in primary tumors. Thus, at this point, we are not able to conclude that the *SLC1A2* 3’ end is involved in another productive fusion. However, 66% of tumors expressing high *SLC1A2* levels did not appear to express *CD44-SLC1A2* (Fig. 3.17). In these non-*CD44-SLC1A2* tumors, it is

possible that *SLC1A2* expression might be driven by fusion to other upstream partners besides *CD44* or due to transcriptional induction of *SLC1A2* by EGF or mTOR/Akt signaling (Wu et al., 2010; Zelenaia et al., 2000). More work will be required to clarify if *SLC1A2* possesses other fusion partners in GC. Similar examples of such "promiscuous" fusion genes have also been reported for *ERG* in prostate cancer and *EWS* in Ewing sarcoma (Sorensen et al., 1994).

To determine the association between 11p13 genomic amplification and *CD44-SLC1A2* fusion expression, we analyzed a panel of fusion positive tumors using Affymetrix SNP6 arrays, which enabled DNA copy number measurements at more than 1.8 million markers across the genome. Of seven fusion-positive cases, two tumors exhibited evidence of 11p13 genomic amplification, while the other five did not (Fig. 3.19). This finding demonstrates that *CD44-SLC1A2* expression may be observed in tumors independently of 11p13 genomic amplification, suggesting that it is not a mere epiphenomenon of 11p13 amplification.

In addition, we investigated the influence of 11p13 amplification and *CD44-SLC1A2* fusion on *CD44* and *SLC1A2* expression. Specifically, we compared *CD44* and *SLC1A2* expression levels across 45 gastric tumors, including a) 11p13 non-amplified samples (32 samples); b) 11p13 amplified but fusion-negative samples (6 samples), and c) *CD44-SLC1A2* fusion positive samples (7 samples) (Fig. 3.20). We noted that the rate of 11p13 amplification in this series (~17%) is similar to frequencies previously reported in the literature (Fukuda et al., 2000). Compared to non-amplified samples, fusion positive

samples exhibited significantly increased *SLC1A2* gene expression, but 11p13 amplified samples did not. In contrast, while *CD44* 3' transcripts were significantly overexpressed in 11p13 amplified tumors, these transcripts were significantly underexpressed in fusion positive tumors (Fig. 3.20). The latter finding is consistent with the *CD44/SLC1A2* genomic inversion decoupling the 3' end of the *CD44* gene body (the region detected by the Affymetrix array) from the endogenous *CD44* promoter. Further evidence of this decoupling was obtained in a real-time PCR analysis measuring *SLC1A2* exon 1, which is not part of the *CD44-SLC1A2* fusion. Unlike the 3' *SLC1A2* transcripts, *SLC1A2* exon 1 was not observed to be highly expressed relative to non-amplified samples. These findings suggest that high *SLC1A2* expression levels are driven more by fusion-events rather than general 11p13 amplification (Fig. 3.20).

In an unsupervised clustering analysis of 197 gastric tumors, we found that a subgroup of *SLC1A2* high-expressing tumors tended to cluster to one another, suggesting that this subgroup of *SLC1A2*-expressing tumors may represent a discrete molecular subclass (Fig. 3.21). Using pathway analysis, we also found that genes upregulated in *SLC1A2*-high expressing tumors were significantly enriched in genes related to ribosomal function and protein translation (Table 3.2), further supporting the notion that *SLC1A2*-high expressing tumors comprise a distinct molecular subclass.

Chapter V: Conclusion

We believe that our study will be of great interest to the GC community, as it represents one of the first recurrent fusion transcripts identified in this disease (and for that matter any common gastrointestinal cancer). Although previous reports have used GBA to identify fusion genes in leukemia and other haematopoietic malignancies, we are the first to show that such a genomic approach can be successfully applied to discover fusion genes in a solid epithelial cancer. In addition, *CD44-SLC1A2* represents a novel example of a fusion gene where the fusion partner (*SLC1A2*) may play a direct role in cancer metabolism, which is different from previously-described fusion genes that have largely involved signalling proteins (kinases) or transcription factors.

As this is the first report that GBA may be successfully applied to discover fusion genes in solid epithelial cancer, such strategy can be used in future to identify fusion genes in other types of solid cancer. In addition, targeting *CD44-SLC1A2* in fusion-positive tumors may also represent a promising avenue for sensitizing GCs to commonly used standard-of-care chemotherapies, since silencing *CD44-SLC1A2* was sufficient to cause a significant sensitization of GC cells to cisplatin *in vitro*. It will be interesting to evaluate the potential of *CD44-SLC1A2* as a drug target. We also showed that patients with tumors expressing high *SLC1A2* levels may represent a distinct molecular subgroup. Thus, *CD44-SLC1A2* and/or high *SLC1A2* levels could be used as a diagnostic biomarker.

As a cell membrane-bound receptor, *CD44-SLC1A2* may prove amenable to targeting using small molecules and reducing intracellular glutamate level may provide to be useful

in treating GC patients. As *SLC1A2* has been extensively studied in Central Nervous System (CNS) associated diseases, and there are numerous glutamate inhibitors available in the market, it would be interesting to test the effects of different glutamate inhibitors to GC cell lines.

Interestingly, we have also identified that several of the *CD44-SLC1A2* positive GCs in our study also exhibited focal amplifications in genes specifically related to RTK/RAS/MAPK signaling, including SNU16 (*FGFR2*), GC2000114 (*MET*), GC2000639 (*KRAS*), and GC980390 (*ERBB2*) (Data not shown). It is thus possible that *CD44-SLC1A2* may collaborate with these canonical oncogenes to facilitate MAPK signalling in gastric cancer cells. Our preliminary study has shown that siRNA specific knockdown of *CD44-SLC1A2* decreased FGFR2 protein level in SNU16 (Data not shown). Taken collectively, one of the future perspectives would be to further investigate the association between *CD44-SLC1A2* and MAPK pathway in GC.

The absolute rate of *CD44-SLC1A2* positivity appeared to be relatively low in this study. On the diagnostic front, we are very interested in validating these findings in larger patient cohorts, both to assess the true frequency of this fusion gene, and also to investigate if there are any clinical or pathologic associations, such as survival rates, associated with *CD44-SLC1A2* positivity.

In addition, 66% of tumors expressing high *SLC1A2* levels did not appear to express *CD44-SLC1A2*. In these non-*CD44-SLC1A2* tumors, it is possible that *SLC1A2*

expression might be driven by fusion to other upstream partners besides *CD44*. More work will be required to clarify if *SLC1A2* possesses other fusion partners in GC.

To our knowledge, *CD44-SLC1A2* represents one of the first recurrent gene fusions identified in a major GI cancer, providing evidence for the existence of this important class of molecular aberrations in GI malignancies. Our results thus strongly warrant additional efforts to identify additional fusion genes in gastrointestinal cancer. For example, transcriptome sequencing approaches have recently identified rare *AGTRAP-BRAF* gene fusions in GC (Palanisamy et al., 2010).

REFERENCES

Akhtar, M., Cheng, Y., Magno, R. M., Ashktorab, H., Smoot, D. T., Meltzer, S. J., and Wilson, K. T. (2001). Promoter methylation regulates Helicobacter pylori-stimulated cyclooxygenase-2 expression in gastric epithelial cells. Cancer Res *61*, 2399-2403.

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., and Sorek, R. (2006). Transcription-mediated gene fusion in the human genome. Genome Res *16*, 30-36.

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988.

Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003). Chromosome aberrations in solid tumors. Nat Genet *34*, 369-376.

Amara, S. G., and Fontana, A. C. (2002). Excitatory amino acid transporters: keeping up with glutamate. Neurochem Int *41*, 313-318.

An, Q., Wright, S. L., Konn, Z. J., Matheson, E., Minto, L., Moorman, A. V., Parker, H., Griffiths, M., Ross, F. M., Davies, T., *et al.* (2008). Variable breakpoints target PAX5 in patients with dicentric chromosomes: a model for the basis of unbalanced translocations in cancer. Proc Natl Acad Sci U S A *105*, 17050-17054.

Aplan, P. D. (2006). Causes of oncogenic chromosomal translocation. Trends Genet *22*, 46-55.

Arriza, J. L., Eliasof, S., Kavanaugh, M. P., and Amara, S. G. (1997). Excitatory amino acid transporter 5, a retinal glutamate transporter coupled to a chloride conductance. Proc Natl Acad Sci U S A *94*, 4155-4160.

Arriza, J. L., Fairman, W. A., Wadiche, J. I., Murdoch, G. H., Kavanaugh, M. P., and Amara, S. G. (1994). Functional comparisons of three glutamate transporter subtypes cloned from human motor cortex. J Neurosci *14*, 5559-5569.

Barber, M., Fitzgerald, R. C., and Caldas, C. (2006). Familial gastric cancer - aetiology and pathogenesis. Best Pract Res Clin Gastroenterol *20*, 721-734.

Beal, M. F. (1992). Mechanisms of excitotoxicity in neurologic diseases. FASEB J *6*, 3338-3344.

Benusiglio, P. R., Pharoah, P. D., Smith, P. L., Lesueur, F., Conroy, D., Luben, R. N., Dew, G., Jordan, C., Dunning, A., Easton, D. F., and Ponder, B. A. (2006). HapMap-based study of the 17q21 ERBB2 amplicon in susceptibility to breast cancer. Br J Cancer *95*, 1689-1695.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., *et al.* (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899-905.

Bignell, G., Santarius, T., Pole, J., Butler, A., Perry, J., Pleasance, E., Greenman, C., Menzies, A., Taylor, S., Edkins, S., *et al.* (2007). Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. Genome Res *17*, 1296-1303.

Block, G., Patterson, B., and Subar, A. (1992). Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. Nutr Cancer *18*, 1-29.

Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K., and Gouaux, E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. Nature *445*, 387-393.

Boussioutas, A., Li, H., Liu, J., Waring, P., Lade, S., Holloway, A. J., Taupin, D., Gorringe, K., Haviv, I., Desmond, P. V., and Bowtell, D. D. (2003). Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. Cancer Res *63*, 2569-2577.

Brenner, H., Rothenbacher, D., and Arndt, V. (2009). Epidemiology of stomach cancer. Methods Mol Biol *472*, 467-477.

Bunch, L., Erichsen, M. N., and Jensen, A. A. (2009). Excitatory amino acid transporters as potential drug targets. Expert Opin Ther Targets *13*, 719-731.

Bärlund, M., Monni, O., Weaver, J. D., Kauraniemi, P., Sauter, G., Heiskanen, M., Kallioniemi, O. P., and Kallioniemi, A. (2002). Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer. Genes Chromosomes Cancer *35*, 311-317.

Canedo, P., Corso, G., Pereira, F., Lunet, N., Suriano, G., Figueiredo, C., Pedrazzani, C., Moreira, H., Barros, H., Carneiro, F., *et al.* (2008). The interferon gamma receptor 1

(IFNGR1) -56C/T gene polymorphism is associated with increased risk of early gastric carcinoma. Gut *57*, 1504-1508.

Cannizzaro, R., and De Paoli, P. (2009). Helicobacter pylori eradication, endoscopic surveillance, and gastric cancer. Am J Gastroenterol *104*, 3100-3101; author reply 3101-3102.

Catalano, V., Labianca, R., Beretta, G. D., Gatta, G., de Braud, F., and Van Cutsem, E. (2009). Gastric cancer. Crit Rev Oncol Hematol *71*, 127-164.

Chuang, L. S., and Ito, Y. (2010). RUNX3 is multifunctional in carcinogenesis of multiple solid tumors. Oncogene *29*, 2605-2615.

Coggon, D., Osmond, C., and Barker, D. J. (1990). Stomach cancer and migration within England and Wales. Br J Cancer *61*, 573-574.

Collins, A. T., Berry, P. A., Hyde, C., Stower, M. J., and Maitland, N. J. (2005). Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res *65*, 10946-10951.

Combaret, V., Gross, N., Lasset, C., Frappaz, D., Peruisseau, G., Philip, T., Beck, D., and Favrot, M. C. (1996). Clinical relevance of CD44 cell-surface expression and N-myc gene amplification in a multicentric analysis of 121 pediatric neuroblastomas. J Clin Oncol *14*, 25-34.

Dalerba, P., Dylla, S. J., Park, I. K., Liu, R., Wang, X., Cho, R. W., Hoey, T., Gurney, A., Huang, E. H., Simeone, D. M., *et al.* (2007). Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A *104*, 10158-10163.

Danbolt, N. C. (2001). Glutamate uptake. Prog Neurobiol *65*, 1-105.

Dang, C. V. (2010). Rethinking the Warburg effect with Myc micromanaging glutamine metabolism. Cancer Res *70*, 859-862.

de Klein, A., van Kessel, A., Grosveld, G., Bartram, C., Hagemeijer, A., Bootsma, D., Spurr, N., Heisterkamp, N., Groffen, J., and Stephenson, J. (1982). A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. Nature *300*, 765-767.

De Marzo, A. M., Bradshaw, C., Sauvageot, J., Epstein, J. I., and Miller, G. J. (1998). CD44 and CD44v6 downregulation in clinical prostatic carcinoma: relation to Gleason grade and cytoarchitecture. Prostate *34*, 162-168.

De Vries, A. C., and Kuipers, E. J. (2007). Review article: Helicobacter pylori eradication for the prevention of gastric cancer. Aliment Pharmacol Ther *26 Suppl 2*, 25-35.

de Vries, A. C., Kuipers, E. J., and Rauws, E. A. (2009). Helicobacter pylori eradication and gastric cancer: when is the horse out of the barn? Am J Gastroenterol *104*, 1342-1345.

DeBerardinis, R. J., and Cheng, T. (2010). Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. Oncogene *29*, 313-324.

Deininger, M., Buchdunger, E., and Druker, B. (2005). The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood *105*, 2640-2653.

Druker, B. J. (2004). Imatinib as a paradigm of targeted therapies. Adv Cancer Res *91*, 1-30.

Druker, B. J., Guilhot, F., O'Brien, S. G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M. W., Silver, R. T., Goldman, J. M., Stone, R. M., *et al.* (2006). Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med *355*, 2408-2417.

Edwards, P. (2010). Fusion genes and chromosome translocations in the common epithelial cancers. J Pathol *220*, 244-254.

El-Omar, E. M., Carrington, M., Chow, W. H., McColl, K. E., Bream, J. H., Young, H. A., Herrera, J., Lissowska, J., Yuan, C. C., Rothman, N., *et al.* (2000). Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature *404*, 398-402.

Espinoza, L. A., Barbieri Neto, J., and Casartelli, C. (1999). Pathological and karyotypic abnormalities in advanced gastric carcinomas. Cancer Genet Cytogenet *109*, 45-50.

Fairman, W. A., Vandenberg, R. J., Arriza, J. L., Kavanaugh, M. P., and Amara, S. G. (1995). An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. Nature *375*, 599-603.

Florijn, R., Bonden, L., Vrolijk, H., Wiegant, J., Vaandrager, J., Baas, F., den Dunnen, J., Tanke, H., van Ommen, G., and Raap, A. (1995). High-resolution DNA Fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. Hum Mol Genet *4*, 831-836.

Fox, J., Beck, P., Dangler, C., and al, e. (2000). Concurrent enteric helminth infection modulates inflammation and gastric immune responses
and reduces helicobacter-induced gastric atrophy. *Nat Med 6*, 536-542.

Fry, L. C., Mönkemüller, K., and Malfertheiner, P. (2007). Prevention of gastric cancer: a challenging but feasible task. Acta Gastroenterol Latinoam *37*, 110-117.

Fukase, K., Kato, M., Kikuchi, S., Inoue, K., Uemura, N., Okamoto, S., Terao, S., Amagai, K., Hayashi, S., Asaka, M., and Group, J. G. S. (2008). Effect of eradication of Helicobacter pylori on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: an open-label, randomised controlled trial. Lancet *372*, 392-397.

Fukuda, Y., Kurihara, N., Imoto, I., Yasui, K., Yoshida, M., Yanagihara, K., Park, J. G., Nakamura, Y., and Inazawa, J. (2000). CD44 is a potential target of amplification within the 11p13 amplicon detected in gastric cancer cell lines. Genes Chromosomes Cancer *29*, 315-324.

G??nthert, U., Stauder, R., Mayer, B., Terpe, H. J., Finke, L., and Friedrichs, K. (1995). Are CD44 variant isoforms involved in human tumour progression? Cancer Surv *24*, 19-42.

Gao, A. C., Lou, W., Dong, J. T., and Isaacs, J. T. (1997). CD44 is a metastasis suppressor gene for prostatic cancer located on human chromosome 11p13. Cancer Res *57*, 846-849.

Gebhardt, F. M., Mitrovic, A. D., Gilbert, D. F., Vandenberg, R. J., Lynch, J. W., and Dodd, P. R. (2010). Exon-skipping splice variants of excitatory amino acid transporter-2 (EAAT2) form heteromeric complexes with full-length EAAT2. J Biol Chem *285*, 31313-31324.

Ghaffari, S., Rostami, S., Bashash, D., Alimoghaddam, K., and Ghavamzadeh, A. (2006). Real-time PCR analysis of PML-RAR alpha in newly diagnosed acute promyelocytic leukaemia patients treated with arsenic trioxide as a front-line therapy. Ann Oncol *17*, 1553-1559.

Gorringe, K. L., Boussioutas, A., Bowtell, D. D., and Melbourne Gastric Cancer Group, P. t. M. M. A. F. (2005). Novel regions of chromosomal amplification at 6p21, 5p13, and 12q14 in gastric cancer identified by array comparative genomic hybridization. Genes Chromosomes Cancer *42*, 247-259.

Graux, C., Cools, J., Melotte, C., Quentmeier, H., Ferrando, A., Levine, R., Vermeesch, J. R., Stul, M., Dutta, B., Boeckx, N., *et al.* (2004). Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. Nat Genet *36*, 1084-1089.

Group, H. a. C. C. (2001). Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut *49*, 347-353.

Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., and Reeve, A. E. (1998). E-cadherin germline mutations in familial gastric cancer. Nature *392*, 402-405.

Guilford, P. J., Hopkins, J. B., Grady, W. M., Markowitz, S. D., Willis, J., Lynch, H., Rajput, A., Wiesner, G. L., Lindor, N. M., Burgart, L. J., *et al.* (1999). E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. Hum Mutat *14*, 249-255.

Hamashima, C., Shibuya, D., Yamazaki, H., Inoue, K., Fukao, A., Saito, H., and Sobue, T. (2008). The Japanese guidelines for gastric cancer screening. Jpn J Clin Oncol *38*, 259-267.

Hartgrink, H., Jansen, E., van Grieken, N., and van de Velde, C. (2009). Gastric cancer. Lancet *374*, 477-490.

Heim, S., and Mitelman, F. (2008). Molecular screening for new fusion genes in cancer. Nat Genet *40*, 685-686.

Herrlich, P., Z??ller, M., Pals, S. T., and Ponta, H. (1993). CD44 splice variants: metastases meet lymphocytes. Immunol Today *14*, 395-399.

Hock, H., Meade, E., Medeiros, S., Schindler, J. W., Valk, P. J., Fujiwara, Y., and Orkin, S. H. (2004). Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. Genes Dev *18*, 2336-2341.

Hohenberger, P., and Gretschel, S. (2003). Gastric cancer. Lancet *362*, 305-315.

Hold, G. L., Rabkin, C. S., Chow, W. H., Smith, M. G., Gammon, M. D., Risch, H. A., Vaughan, T. L., McColl, K. E., Lissowska, J., Zatonski, W., *et al.* (2007). A functional polymorphism of toll-like receptor 4 gene increases risk of gastric carcinoma and its precursors. Gastroenterology *132*, 905-912.

Hou, Q., Wu, Y., Grabsch, H., Zhu, Y., Leong, S., Ganesan, K., Cross, D., Tan, L., Tao, J., Gopalakrishnan, V., *et al.* (2008). Integrative genomics identifies RAB23 as an invasion mediator gene in diffuse-type gastric cancer. Cancer Res *68*, 4623-4630.

Hughes, T. P., Kaeda, J., Branford, S., Rudzki, Z., Hochhaus, A., Hensley, M. L., Gathmann, I., Bolton, A. E., van Hoomissen, I. C., Goldman, J. M., *et al.* (2003). Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. N Engl J Med *349*, 1423-1432.

Husdal, A., Bukholm, G., and Bukholm, I. R. (2006). The prognostic value and overexpression of cyclin A is correlated with gene amplification of both cyclin A and cyclin E in breast cancer patient. Cell Oncol *28*, 107-116.

Imazeki, F., Yokosuka, O., Yamaguchi, T., Ohto, M., Isono, K., and Omata, M. (1996). Expression of variant CD44-messenger RNA in colorectal adenocarcinomas and adenomatous polyps in humans. Gastroenterology *110*, 362-368.

International Agency for Research on Cancer, W. H. O. (1994). Infection with *Helicobacter pylori* , in *Schistosomes,liver flukes and Helicobacter pylori*. In, (IARC Monogr. Eval. Carcinog. Risks Hum.), pp. 177-240.

Ishimoto, T., Nagano, O., Yae, T., Tamada, M., Motohara, T., Oshima, H., Oshima, M., Ikeda, T., Asaba, R., Yagi, H., *et al.* (2011). CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc(-) and thereby promotes tumor growth. Cancer Cell *19*, 387-400.

Ishimoto, T., Oshima, H., Oshima, M., Kai, K., Torii, R., Masuko, T., Baba, H., Saya, H., and Nagano, O. (2010). CD44+ slow-cycling tumor cell expansion is triggered by cooperative actions of Wnt and prostaglandin E2 in gastric tumorigenesis. Cancer Sci *101*, 673-678.

Jackson, C., Cunningham, D., Oliveira, J., and Group, E. G. W. (2009). Gastric cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol *20 Suppl 4*, 34-36.

Kallioniemi, A. (2008). CGH microarrays and cancer. Curr Opin Biotechnol *19*, 36-40.

Kamangar, F., Dores, G., and Anderson, W. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol *24*, 2137-2150.

Kanai, Y., and Hediger, M. A. (1992). Primary structure and functional characterization of a high-affinity glutamate transporter. Nature *360*, 467-471.

Kang, J. U., Kang, J. J., Kwon, K. C., Park, J. W., Jeong, T. E., Noh, S. M., and Koo, S. H. (2006). Genetic alterations in primary gastric carcinomas correlated with clinicopathological variables by array comparative genomic hybridization. J Korean Med Sci *21*, 656-665.

Kawamata, N., Ogawa, S., Zimmermann, M., Niebuhr, B., Stocking, C., Sanada, M., Hemminki, K., Yamatomo, G., Nannya, Y., Koehler, R., *et al.* (2008). Cloning of genes

involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray. Proc Natl Acad Sci U S A *105*, 11921-11926.

Kim, J. Y., Bae, B. N., Kim, K. S., Shin, E., and Park, K. (2009). Osteopontin, CD44, and NFkappaB expression in gastric adenocarcinoma. Cancer Res Treat *41*, 29-35.

Kimura, Y., Noguchi, T., Kawahara, K., Kashima, K., Daa, T., and Yokoyama, S. (2004). Genetic alterations in 102 primary gastric cancers by comparative genomic hybridization: gain of 20q and loss of 18q are associated with tumor progression. Mod Pathol *17*, 1328-1337.

Kono, S., and Hirohata, T. (1996). Nutition and stomach cancer. . *Cancer Causes Control 7*, 41-55.

Krejs, G. J. (2010). Gastric cancer: epidemiology and risk factors. Dig Dis *28*, 600-603.

Kroemer, G., and Pouyssegur, J. (2008). Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell *13*, 472-482.

Kruse, J. P., and Gu, W. (2009). Modes of p53 regulation. Cell *137*, 609-622.

Kumar-Sinha, C., Tomlins, S. A., and Chinnaiyan, A. M. (2008). Recurrent gene fusions in prostate cancer. Nat Rev Cancer *8*, 497-511.

LAG, R., D, M., M, K., and *al*., *e.* (2008). *SEER Cancer Statistics Review, 1975-2005*. In, (National Cancer Institute, Bethesda, MD,).

Lauren, P. (1965). The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. Acta Pathol Microbiol Scand *64*, 31-49.

Lauriat, T. L., and McInnes, L. A. (2007). EAAT2 regulation and splicing: relevance to psychiatric and neurological disorders. Mol Psychiatry *12*, 1065-1078.

Lee, W. C. (2006). Breast, stomach and colorectal cancer screening in Korea. J Med Screen *13 Suppl 1*, S20-22.

Li, Q., Ito, K., Sakakura, C., Fukamachi, H., Inoue, K., Chi, X., Lee, K., Nomura, S., Lee, C., Han, S., *et al.* (2002). Causal relationship between the loss of RUNX3 expression and gastric cancer. Cell *109*, 113-124.

Li, X., and Francke, U. (1995). Assignment of the gene SLC1A2 coding for the human glutamate transporter EAAT2 to human chromosome 11 bands p13-p12. Cytogenet Cell Genet *71*, 212-213.

Lim, G. H., Wong, C. S., Chow, K. Y., Bhalla, V., and Chia, K. S. (2009). Trends in long-term cancer survival in Singapore: 1968-2002. Ann Acad Med Singapore *38*, 99-105.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B. G., Ohgi, K., Zhang, J., Rose, D. W., Fu, X. D., Glass, C. K., and Rosenfeld, M. G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

Look, A. T. (1997). Oncogenic transcription factors in the human acute leukemias. Science *278*, 1059-1064.

Lu, W., Pelicano, H., and Huang, P. (2010). Cancer metabolism: is glutamine sweeter than glucose? Cancer Cell *18*, 199-200.

Maher, C., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A. (2009). Transcriptome sequencing to detect gene fusions in cancer. Nature *458*, 97-101.

Malfertheiner, P., Sipponen, P., Naumann, M., Moayyedi, P., Mégraud, F., Xiao, S. D., Sugano, K., Nyrén, O., and Force, L. H. p.-G. C. T. (2005). Helicobacter pylori eradication has the potential to prevent gastric cancer: a state-of-the-art critique. Am J Gastroenterol *100*, 2100-2115.

Mani, R. S., Tomlins, S. A., Callahan, K., Ghosh, A., Nyati, M. K., Varambally, S., Palanisamy, N., and Chinnaiyan, A. M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. Science *326*, 1230.

Marcaggi, P., and Attwell, D. (2004). Role of glial amino acid transporters in synaptic transmission and brain energetics. Glia *47*, 217-225.

Marshall, B. J., and Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet *1*, 1311-1315.

Medina, M., Sánchez-Jiménez, F., Márquez, J., Rodríguez Quesada, A., and Núñez de Castro, I. (1992). Relevance of glutamine metabolism to tumor cell growth. Mol Cell Biochem *113*, 1-15.

Meyer, T., Ludolph, A. C., Morkel, M., Hagemeier, C., and Speer, A. (1997). Genomic organization of the human excitatory amino acid transporter gene GLT-1. Neuroreport *8*, 775-777.

Meyer, T., Munch, C., Knappenberger, B., Liebau, S., Volkel, H., and Ludolph, A. C. (1998). Alternative splicing of the glutamate transporter EAAT2 (GLT-1). Neurosci Lett *241*, 68-70.

Miki, K. (2006). Gastric cancer screening using the serum pepsinogen test method. Gastric Cancer *9*, 245-253.

Mitelman, F. (2000). Recurrent chromosome aberrations in cancer. Mutat Res *462*, 247-253.

Mitelman, F., Johansson, B., and Mertens, F. (2004). Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet *36*, 331-334.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer *7*, 233-245.

Mitelman, F., Mertens, F., and Johansson, B. (2005). Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders. Genes Chromosomes Cancer *43*, 350-366.

Mitsui, F., Dobashi, Y., Imoto, I., Inazawa, J., Kono, K., Fujii, H., and Ooi, A. (2007). Non-incidental coamplification of Myc and ERBB2, and Myc and EGFR, in gastric adenocarcinomas. Mod Pathol *20*, 622-631.

Morris, D. S., Tomlins, S. A., Montie, J. E., and Chinnaiyan, A. M. (2008). The discovery and application of gene fusions in prostate cancer. BJU Int *102*, 276-282.

Mullighan, C., Goorha, S., Radtke, I., Miller, C., Coustan-Smith, E., Dalton, J., Girtman, K., Mathew, S., Ma, J., Pounds, S., *et al.* (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature *446*, 758-764.

Munch, C., Schwalenstocker, B., Knappenberger, B., Liebau, S., Volkel, H., Ludolph, A. C., and Meyer, T. (1998). 5'-heterogeneity of the human excitatory amino acid transporter cDNA EAAT2 (GLT-1). Neuroreport *9*, 1295-1297.

Nagano, O., and Saya, H. (2004). Mechanism and biological significance of CD44 cleavage. Cancer Sci *95*, 930-935.

Nakanishi, M., Sakakura, C., Fujita, Y., Yasuoka, R., Aragane, H., Koide, K., Hagiwara, A., Yamaguchi, T., Nakamura, Y., Abe, T., *et al.* (2000). Genomic alterations in primary gastric cancers analyzed by comparative genomic hybridization and clinicopathological factors. Hepatogastroenterology *47*, 658-662.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D. K., Kennedy, G. C., and Ogawa, S. (2005). A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res *65*, 6071-6079.

Nessling, M., Solinas-Toldo, S., Wilgenbus, K. K., Borchard, F., and Lichter, P. (1998). Mapping of chromosomal imbalances in gastric adenocarcinoma revealed amplified protooncogenes MYCN, MET, WNT2, and ERBB2. Genes Chromosomes Cancer *23*, 307-316.

Neugut, A. I., Hayek, M., and Howe, G. (1996). Epidemiology of gastric cancer. Semin Oncol *23*, 281-291.

Nicklin, P., Bergman, P., Zhang, B., Triantafellow, E., Wang, H., Nyfeler, B., Yang, H., Hild, M., Kung, C., Wilson, C., *et al.* (2009). Bidirectional transport of amino acids regulates mTOR and autophagy. Cell *136*, 521-534.

Niini, T., López-Guerrero, J., Ninomiya, S., Guled, M., Hattinger, C., Michelacci, F., Böhling, T., Llombart-Bosch, A., Picci, P., Serra, M., and Knuutila, S. (2010). Frequent deletion of CDKN2A and recurrent coamplification of KIT, PDGFRA, and KDR in fibrosarcoma of bone--an array comparative genomic hybridization study. Genes Chromosomes Cancer *49*, 132-143.

Nowell, P. C., and Hungerford, D. A. (1960). A minute chromosome in human chronic granulocytic leukemia. Science *132*, 1497.

O'Brien, S. G., Guilhot, F., Larson, R. A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J. J., Fischer, T., Hochhaus, A., Hughes, T., *et al.* (2003). Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med *348*, 994-1004.

Okada, A., Takehara, H., Yoshida, K., Nishi, M., Miyake, H., Kita, Y., and Komi, N. (1993). Increased aspartate and glutamate levels in both gastric and colon cancer tissues. Tokushima J Exp Med *40*, 19-25.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G., and Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.

Ooi, C., Ivanova, T., Wu, J., Lee, M., Tan, I., Tao, J., Ward, L., Koo, J., Gopalakrishnan, V., Zhu, Y., *et al.* (2009). Oncogenic pathway combinations predict clinical prognosis in gastric cancer. PLoS Genet *5*, e1000676.

Orian-Rousseau, V., Chen, L., Sleeman, J. P., Herrlich, P., and Ponta, H. (2002). CD44 is required for two consecutive steps in HGF/c-Met signaling. Genes Dev *16*, 3074-3086.

Palanisamy, N., Ateeq, B., Kalyana-Sundaram, S., Pflueger, D., Ramnarayanan, K., Shankar, S., Han, B., Cao, Q., Cao, X., Suleman, K., *et al.* (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med *16*, 793-798.

Palli, D. (2000). Epidemiology of gastric cancer: an evaluation of available evidence. J Gastroenterol *35 Suppl 12*, 84-89.

Pan, K., Liu, W., Zhang, L., You, W., and Lu, Y. (2008). Mutations in components of the Wnt signaling pathway in gastric cancer. World J Gastroenterol *14*, 1570-1574.

Park, J. G., Frucht, H., LaRocca, R. V., Bliss, D. P., Jr., Kurita, Y., Chen, T. R., Henslee, J. G., Trepel, J. B., Jensen, R. T., Johnson, B. E., and et al. (1990). Characteristics of cell lines established from human gastric carcinoma. Cancer Res *50*, 2773-2780.

Parkin, D. M., Bray, F., Ferlay, J., and Pisani, P. (2005). Global cancer statistics, 2002. CA Cancer J Clin *55*, 74-108.

Persson, F., Andrén, Y., Winnes, M., Wedell, B., Nordkvist, A., Gudnadottir, G., Dahlenfors, R., Sjögren, H., Mark, J., and Stenman, G. (2009). High-resolution genomic profiling of adenomas and carcinomas of the salivary glands reveals amplification, rearrangement, and fusion of HMGA2. Genes Chromosomes Cancer *48*, 69-82.

Pines, G., Danbolt, N. C., Bjoras, M., Zhang, Y., Bendahan, A., Eide, L., Koepsell, H., Storm-Mathisen, J., Seeberg, E., and Kanner, B. I. (1992). Cloning and expression of a rat brain L-glutamate transporter. Nature *360*, 464-467.

Ponta, H., Sherman, L., and Herrlich, P. A. (2003). CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol *4*, 33-45.

Pui, C. H., Relling, M. V., and Downing, J. R. (2004). Acute lymphoblastic leukemia. N Engl J Med *350*, 1535-1548.

Pui, C. H., Robison, L. L., and Look, A. T. (2008). Acute lymphoblastic leukaemia. Lancet *371*, 1030-1043.

Rabbitts, T., and Boehm, T. (1991). Structural and functional chimerism results from chromosomal translocation in lymphoid tumors. Adv Immunol *50*, 119-146.

Rabbitts, T. H. (1994). Chromosomal translocations in human cancer. Nature *372*, 143-149.

Rastogi, T., Hildesheim, A., and Sinha, R. (2004). Opportunities for cancer epidemiology in developing countries. Nat Rev Cancer *4*, 909-917.

Ren, R. (2005). Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat Rev Cancer *5*, 172-183.

Riedel, G., Platt, B., and Micheau, J. (2003). Glutamate receptor function in learning and memory. Behav Brain Res *140*, 1-47.

Rodriguez, E., Rao, P. H., Ladanyi, M., Altorki, N., Albino, A. P., Kelsen, D. P., Jhanwar, S. C., and Chaganti, R. S. (1990). 11p13-15 is a specific region of chromosomal rearrangement in gastric and esophageal adenocarcinomas. Cancer Res *50*, 6410-6416.

Rothenbacher, D., and Brenner, H. (2003). Burden of Helicobacter pylori and H. pylori-related diseases in developed countries: recent developments and future implications. Microbes Infect *5*, 693-703.

Rothstein, J. D., Dykes-Hoberg, M., Pardo, C. A., Bristol, L. A., Jin, L., Kuncl, R. W., Kanai, Y., Hediger, M. A., Wang, Y., Schielke, J. P., and Welty, D. F. (1996). Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate. Neuron *16*, 675-686.

Rowley, J. D., and Blumenthal, T. (2008). Medicine. The cart before the horse. Science *321*, 1302-1304.

Rubin, M., and Chinnaiyan, A. (2006). Bioinformatics approach leads to the discovery of the TMPRSS2:ETS gene fusion in prostate cancer. Lab Invest *86*, 1099-1102.

Rzeski, W., Ikonomidou, C., and Turski, L. (2002). Glutamate antagonists limit tumor growth. Biochem Pharmacol *64*, 1195-1200.

Rzeski, W., Turski, L., and Ikonomidou, C. (2001). Glutamate antagonists limit tumor growth. Proc Natl Acad Sci U S A *98*, 6372-6377.

Sakakura, C., Mori, T., Sakabe, T., Ariyama, Y., Shinomiya, T., Date, K., Hagiwara, A., Yamaguchi, T., Takahashi, T., Nakamura, Y., *et al.* (1999). Gains, losses, and

amplifications of genomic materials in primary gastric cancers analyzed by comparative genomic hybridization. Genes Chromosomes Cancer *24*, 299-305.

Sastre, J., Garcia-Saenz, J. A., and Diaz-Rubio, E. (2006). Chemotherapy for gastric cancer. World J Gastroenterol *12*, 204-213.

Sato, K., Tamura, G., Tsuchiya, T., Endoh, Y., Sakata, K., Motoyama, T., Usuba, O., Kimura, W., Terashima, M., Nishizuka, S., *et al.* (2002). Analysis of genetic and epigenetic alterations of the PTEN gene in gastric cancer. Virchows Arch *440*, 160-165.

Schmits, R., Filmus, J., Gerwin, N., Senaldi, G., Kiefer, F., Kundig, T., Wakeham, A., Shahinian, A., Catzavelos, C., Rak, J., *et al.* (1997). CD44 regulates hematopoietic progenitor distribution, granuloma formation, and tumorigenicity. Blood *90*, 2217-2233.

Screaton, G. R., Bell, M. V., Bell, J. I., and Jackson, D. G. (1993). The identification of a new alternative exon with highly restricted tissue expression in transcripts encoding the mouse Pgp-1 (CD44) homing receptor. Comparison of all 10 variable exons between mouse, human, and rat. J Biol Chem *268*, 12235-12238.

Selgrad, M., Bornschein, J., Rokkas, T., and Malfertheiner, P. (2010). Clinical aspects of gastric cancer and Helicobacter pylori--screening, prevention, and treatment. Helicobacter *15 Suppl 1*, 40-45.

Sen, S., Zhou, H., and White, R. A. (1997). A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. Oncogene *14*, 2195-2200.

Shtivelman, E., and Bishop, J. M. (1991). Expression of CD44 is repressed in neuroblastoma cells. Mol Cell Biol *11*, 5446-5453.

Sims, K. D., and Robinson, M. B. (1999). Expression patterns and regulation of glutamate transporters in the developing and adult nervous system. Crit Rev Neurobiol *13*, 169-197.

Sirvent, N., Maire, G., and Pedeutour, F. (2003). Genetics of dermatofibrosarcoma protuberans family of tumors: from ring chromosomes to tyrosine kinase inhibitor treatment. Genes Chromosomes Cancer *37*, 1-19.

Soda, M., Choi, Y., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., *et al.* (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature *448*, 561-566.

Sorensen, P. H., Lessnick, S. L., Lopez-Terrada, D., Liu, X. F., Triche, T. J., and Denny, C. T. (1994). A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. Nat Genet *6*, 146-151.

Soussi, T., Dehouche, K., and Béroud, C. (2000). p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. Hum Mutat *15*, 105-113.

Stamenkovic, I., Aruffo, A., Amiot, M., and Seed, B. (1991). The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. EMBO J *10*, 343-348.

Stauder, R., Eisterer, W., Thaler, J., and G??nthert, U. (1995). CD44 variant isoforms in non-Hodgkin's lymphoma: a new independent prognostic factor. Blood *85*, 2885-2899.

Stemmermann, G. N. (1994). Intestinal metaplasia of the stomach. A status report. Cancer *74*, 556-564.

Storck, T., Schulte, S., Hofmann, K., and Stoffel, W. (1992). Structure, expression, and functional analysis of a Na(+)-dependent glutamate/aspartate transporter from rat brain. Proc Natl Acad Sci U S A *89*, 10955-10959.

Su, Z. Z., Leszczyniecka, M., Kang, D. C., Sarkar, D., Chao, W., Volsky, D. J., and Fisher, P. B. (2003). Insights into glutamate transport regulation in human astrocytes: cloning of the promoter for excitatory amino acid transporter 2 (EAAT2). Proc Natl Acad Sci U S A *100*, 1955-1960.

Tahara, E. (1995). Molecular biology of gastric cancer. World J Surg *19*, 484-488; discussion 489-490.

Takano, T., Lin, J., Arcuino, G., Gao, Q., Yang, J., and Nedergaard, M. (2001). Glutamate release promotes growth of malignant gliomas. Nat Med *7*, 1010-1015.

Tamura, G., Kihana, T., Nomura, K., Terada, M., Sugimura, T., and Hirohashi, S. (1991). Detection of frequent p53 gene mutations in primary gastric cancer by cell sorting and polymerase chain reaction single-strand conformation polymorphism analysis. Cancer Res *51*, 3056-3058.

Tanabe, K. K., Ellis, L. M., and Saya, H. (1993). Expression of CD44R1 adhesion molecule in colon carcinomas and metastases. Lancet *341*, 725-726.

Tay, S., Leong, S., Yu, K., Aggarwal, A., Tan, S., Lee, C., Wong, K., Visvanathan, J., Lim, D., Wong, W., *et al.* (2003). A combined comparative genomic hybridization and

expression microarray analysis of gastric cancer reveals novel molecular subtypes. Cancer Res *63*, 3309-3316.

Tolg, C., Hofmann, M., Herrlich, P., and Ponta, H. (1993). Splicing choice from ten variant exons establishes CD44 variability. Nucleic Acids Res *21*, 1225-1229.

Tomlins, S., Rhodes, D., Perner, S., Dhanasekaran, S., Mehra, R., Sun, X., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644-648.

Tsukamoto, Y., Uchida, T., Karnan, S., Noguchi, T., Nguyen, L. T., Tanigawa, M., Takeuchi, I., Matsuura, K., Hijiya, N., Nakada, C., *et al.* (2008). Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer. J Pathol *216*, 471-482.

Tsukita, S., Oishi, K., Sato, N., Sagara, J., and Kawai, A. (1994). ERM family members as molecular linkers between the cell surface glycoprotein CD44 and actin-based cytoskeletons. J Cell Biol *126*, 391-401.

Vaananen, H., Vauhkonen, M., Helske, T., Kaariainen, I., Rasmussen, M., Tunturi-Hihnala, H., Koskenpato, J., Sotka, M., Turunen, M., Sandstrom, R., *et al.* (2003). Non-endoscopic diagnosis of atrophic gastritis with a blood test. Correlation between gastric histology and serum levels of gastrin-17 and pepsinogen I: a multicentre study. Eur J Gastroenterol Hepatol *15*, 885-891.

Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029-1033.

Varis, A., Wolf, M., Monni, O., Vakkari, M., Kokkola, A., Moskaluk, C., Frierson, H. J., Powell, S., Knuutila, S., Kallioniemi, A., and El-Rifai, W. (2002). Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res *62*, 2625-2629.

Wagner, A. D., Grothe, W., Haerting, J., Kleber, G., Grothey, A., and Fleig, W. E. (2006). Chemotherapy in advanced gastric cancer: a systematic review and meta-analysis based on aggregate data. J Clin Oncol *24*, 2903-2909.

Wang, K., Ubriaco, G., and Sutherland, L. C. (2007). RBM6-RBM5 transcription-induced chimeras are differentially expressed in tumours. BMC Genomics *8*, 348.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H., and Speck, N. A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci U S A *93*, 3444-3449.

Watabe, H., Mitsushima, T., Yamaji, Y., Okamoto, M., Wada, R., Kokubo, T., Doi, H., Yoshida, H., Kawabe, T., and Omata, M. (2005). Predicting the development of gastric cancer from combining Helicobacter pylori antibodies and serum pepsinogen status: a prospective endoscopic cohort study. Gut *54*, 764-768.

WCRF, and AICR (2007). Food, nutrition, physical activity and the prevention of cancer: a global perspective. In, (Washington, DC: AICR).

Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjöblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., *et al.* (2007). The genomic landscapes of human breast and colorectal cancers. Science *318*, 1108-1113.

Wu, C. Y., Kuo, K. N., Wu, M. S., Chen, Y. J., Wang, C. B., and Lin, J. T. (2009). Early Helicobacter pylori eradication decreases risk of gastric cancer in patients with peptic ulcer disease. Gastroenterology *137*, 1641-1648.e1641-1642.

Wu, X., Kihara, T., Akaike, A., Niidome, T., and Sugimoto, H. (2010). PI3K/Akt/mTOR signaling regulates glutamate transporter 1 in astrocytes. Biochem Biophys Res Commun *393*, 514-518.

Yang, C., Sudderth, J., Dang, T., Bachoo, R., Bachoo, R., McDonald, J., and DeBerardinis, R. (2009). Glioblastoma cells require glutamate dehydrogenase to survive impairments of glucose metabolism or Akt signaling. Cancer Res *69*, 7986-7993.

Ye, R., Rhoderick, J. F., Thompson, C. M., and Bridges, R. J. (2010). Functional expression, purification and high sequence coverage mass spectrometric characterization of human excitatory amino acid transporter EAAT2. Protein Expr Purif *74*, 49-59.

Zelenaia, O., Schlag, B. D., Gochenauer, G. E., Ganel, R., Song, W., Beesley, J. S., Grinspan, J. B., Rothstein, J. D., and Robinson, M. B. (2000). Epidermal growth factor receptor agonists increase expression of glutamate transporter GLT-1 in astrocytes through pathways dependent on phosphatidylinositol 3-kinase and transcription factor NF-kappaB. Mol Pharmacol *57*, 667-678.